UNIVERSITY OF LEEDS

This is a repository copy of *Path Size Logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/156992/

Version: Accepted Version

Article:
Duncan, LC, Watling, DP orcid.org/0000-0002-6193-9121, Connors, RD orcid.org/0000-0002-1696-0175 et al. (2 more authors) (2020) Path Size Logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. Transportation Research Part B: Methodological, 135. pp. 1-40. ISSN 0191-2615

https://doi.org/10.1016/j.trb.2020.02.006

© 2020 Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Path Size Logit Route Choice Models: Issues with Current Models, a New Internally Consistent Approach, and Parameter Estimation on a Large-Scale Network with GPS Data

Lawrence Christopher DUNCAN [a] *, David Paul WATLING [a], Richard Dominic CONNORS [a], Thomas Kjær RASMUSSEN [b], Otto Anker NIELSEN [b]

[a] Institute for Transport Studies, University of Leeds
36-40 University Road, Leeds, LS2 9JT, United Kingdom.
[b] Department of Technology, Management and Economics, Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark.

* corresponding author:
Institute for Transport Studies, University of Leeds
36-40 University Road, Leeds, LS2 9JT, United Kingdom
Tel.: +441133436612
E-mail: ts16ld@leeds.ac.uk

Highlights

- Demonstrate issues with existing Path Size Logit (PSL) models
- Propose a new internally consistent Adaptive PSL that addresses these issues
- Proof of existence and uniqueness conditions for Adaptive PSL solutions
- Adaptive PSL Likelihood formulation and tracked route data MLE procedure
- Estimation of PSL models on a large-scale network using real GPS data

Abstract

Path Size Logit route choice models attempt to capture the correlation between routes by including correction terms within the route utility functions. This provides a convenient closed-form solution for implementation in traffic network models. The path size terms measure distinctiveness of routes; a route is penalised based on the number of other routes sharing its links, and the costs of those shared links. Typically, real road networks have many very long routes that should be considered unrealistic. Such unrealistic routes are problematic for the Path Size Logit (PSL) model because they negatively impact the choice probabilities of realistic routes when links are shared. The Generalised Path Size Logit (GPSL) model attempts to address this problem by weighting the contributions of routes to path size terms according to the ratio of route travel costs. However, the GPSL model is not internally consistent in how it defines routes as being unrealistic: the path size terms consider only travel cost, whereas the route choice probability relation considers disutility *including* the correction term.

To solve these challenges, this paper formulates a new internally consistent Adaptive Path Size Logit (APSL) model wherein routes contribute to path size terms according to the ratio of route choice probabilities, ensuring that routes defined as unrealistic by the path size terms, are exactly those with very low choice probabilities. The APSL route choice probability relation is an implicit function, naturally expressed as a fixed-point problem. A proof is provided for the guaranteed existence of solutions, as well as conditions for the uniqueness of solutions. A Maximum Likelihood Estimation procedure is given for estimating the APSL model with tracked route observation data, and this procedure is investigated in a simulation study where it is shown it is generally possible to reproduce assumed true parameters. APSL is then estimated using real tracked route GPS data on a large-scale network, and results are compared with other PSL models.

Key Words: path size logit, route choice, random utility, fixed-point problem, overlapping routes, parameter estimation

1 Introduction

It is well known that the Multinomial Logit (MNL) Random Utility Model (RUM) often provides unrealistic choice probabilities when applied to real road network route choice. One of the main reasons for this is that the MNL model does not capture the correlation between routes. This issue stems from the underlying assumption made by the MNL model that the random error terms are independently and identically distributed (IID) with the same, fixed variances (Sheffi, 1985). Numerous adaptations of the MNL model have been proposed in the literature which relax the IID assumption – specifically the independently distributed assumption – and attempt to capture the correlation between the

routes. These extended Logit models can be classified into three groups according to the model structures as suggested by Prashker & Bekhor (2004): GEV structures, Mixed Logit models, and MNL-modification models. The fourth group of route choice models that provide an option for overcoming the weakness of MNL are those which propose an alternative probability distribution for the random error terms, i.e. alternative RUMs, that capture the correlation implicitly. To set the background for the research in the present paper, we consider each of these categories in turn below:

GEV structures: The first group of extended Logit models are those which are based on the Generalized Extreme Value (GEV) theory (McFadden, 1978), which use a two-level tree structure to capture the similarity among routes through the random error component of the utility function. These models include the Cross-Nested Logit (CNL) model (Vovsha, 1997; Bekhor & Prashker, 1999; Marzano & Papola, 2008), the Paired Combinatorial Logit (PCL) model (Chu, 1989; Bekhor & Prashker, 1999; Gliebe et al, 1999; Pravinvongvuth & Chen, 2005), the Generalized Nested Logit (GNL) model (Bekhor & Prashker, 2001; Wen & Koppelman, 2001), and the Network GEV model (Bierlaire, 2002; Daly & Bierlaire, 2006) which uses a fairly general class of networks to generalise the use of trees to represent nested logit models to a network representation. The PCL, CNL, and GNL models all have closed-form probability expressions, though due to their two-level tree structure the route choice probabilities are complex to compute for large-scale network applications. Furthermore, several studies have found that when estimating the nesting coefficients for the CNL model, the model tends to collapse to MNL (Ramming, 2002; Prato, 2005; Prato & Bekhor, 2006), and the GNL model requires the estimation of an additional parameter over the CNL model which makes parameter estimation more difficult. The network GEV model

Mixed Logit models: The second group of extended Logit models are Mixed Logit models (Ben-Akiva & Bolduc, 1996; McFadden & Train, 2000), also known as Logit Kernel, Random Parameter Logit, Error Component Logit, and Hybrid Logit. Mixed Logit models attempt to capture the correlation between routes by dividing the random error terms into two components; the first component is a set of IID Gumbel variables ensuring that the Logit structure is kept, and the second component is a set of Gaussian distributed variable terms that attempt to capture the interdependencies among the routes. Bekhor et al (2002) propose a Factor Analytic Logit Kernel model that attempts to capture the similarities among routes by assuming the covariance between utilities relates to overlap lengths. The main drawback of Mixed Logit models however, is that they do not possess closed-form expressions and therefore solving the route choice probabilities requires either Monte Carlo simulation or similar methods, which are computationally burdensome. There are also difficulties in estimating the parameters of the Factor Analytic Logit Kernel model: Ramming (2002) finds instable estimates of the covariance parameters, while Prato (2005) discusses the difficulty in obtaining significant estimates.

MNL-modification models: The third group of extended Logit models are the MNL-modification models which modify the deterministic part of the route utilities by including a correction term that adjusts the route choice probabilities to approximate the correlation between the routes. These models include C-Logit (CL) (Cascetta et al, 1996), Path Size Logit (PSL) (Ben-Akiva & Ramming, 1998), and Path Size Correction Logit (PSCL) (Bovy et al, 2008). Ramming (2002) proposes a Generalised Path Size Logit (GPSL) model which includes a component that attempts to reduce the impact that infeasibly long routes have on the correction terms (and thus choice probabilities) of feasible routes. The main attraction of MNL-modification models is that they all retain the single-level tree structure as MNL and have simple closed-form expressions, meaning the route choice probabilities are generally easy and quick to compute, and estimating the parameters of the CL, PSL, and PSCL models is a comparatively simple task, though as the size of network increases, so does the computational effort required to enumerate the overlap between all routes in a choice set. The GPSL model requires the estimation of an additional parameter over the PSL model which makes parameter estimation more difficult; estimates for the parameter can be justified by assessing the goodness-of-fit, though the best estimates in case studies tend to be very large values meaning that it is difficult to provide a behavioural interpretation for the parameter.

Alternative RUMs: Another option is to utilise an alternative RUM. There are many RUMs that do not suffer from the same issue as MNL as the similarity between each pair of routes is accounted for by allowing for covariance between the error terms. The MNL RUM proposes that the random error terms assume a Gumbel distribution (Dial, 1971), while the Multinomial Probit (MNP) model (Daganzo & Sheffi, 1977) and the Multinomial Gammit (MNG) model (Cantarella & Binetti, 2002), propose that the error terms assume a Normal distribution and Gamma distribution, respectively. Other distributions include Log-Normal and Uniform, while Nielsen (2000) argues that Gamma distributed error terms is preferred since the error term is positive and link-additive (Nielsen & Frederiksen, 2006). These models however, do not have closed-form probability expressions and hence solving the route choice probabilities also requires either Monte Carlo simulation or similar methods which are computationally burdensome and converge very slowly on large scale networks (Rich et al, 2015; Manzo et al, 2015; Rasmussen et al, 2016; Connors et al, 2014). Mishra et al (2012), Ahipasaoglu et al (2013), and Ahipasaoglu et al (2015) explore a Cross Moment (CMM) choice model where the exact distribution of the random error terms is unknown and instead belongs to a set, where the distribution employed is that

which maximises the expected utility given known mean and covariance information of the utilities. Efficient convex optimisation techniques are developed to solve the CMM, though computation times increase dramatically as the number of routes increases.

There are numerous examples of alternative RUMs where the correlation between routes is not explicitly captured (like MNL) and have been adapted accordingly utilising concepts from extended Logit models. These models thus share the same associated strengths/weaknesses of the approaches. Castillo et al (2008) proposed a Multinomial Weibit (MNW) model – where the random error terms assume a Weibull distribution – to address the other underlying assumption made by the MNL model that the error terms are identically distributed. Kitthamkesorn & Chen (2013) then integrated the ideas of the MNW and PSL models to formulate a Path Size Weibit (PSW) model that simultaneously addresses both the independently distributed and identically distributed assumptions made of the MNL error terms. Xu et al (2015) formulate a Hybrid closed-form route choice model to alleviate the contrasting scaling issues of MNL and MNW by simultaneously considering absolute cost difference and relative cost difference, and is extended to include a path size correction factor to capture the correlation between routes. Chikaraishi & Nakayama (2016) extend concepts from the q-Generalised Logit model (Nakayama & Chikaraishi, 2015) to introduce a q-Product Logit model in which the relationship between the deterministic and random components of utilities can be either additive, multiplicative, or in-between, depending on the value of the parameter q, where MNL and MNW are special cases of the model. A q-Product Nested Logit model is presented to capture correlation, where the CNL model is a special case, as well as a nested equivalent of the Weibit model. Li (2011) proposes a Semi-Parametric choice model that relaxes the assumption of underlying distributions from either Gumbel or Weibull to a wider distribution class where the underlying choice model is unknown, and integrates Mixed Logit concepts to postulate a Mixed Semi-Parametric choice model. Ahipasaoglu et al (2016) consider the application of the Marginal Distribution Model (MDM) (Natarajan et al, 2009) to route choice, where the marginal distributions of the route utilities are specified but the joint distributions are not, and the focus is on the particular joint distribution that maximizes expected utility. Incorporating information on the marginal distributions makes the MDM model flexible and MNL, CL, PSL, MNW, and PSW are all special cases. Numerous variants of the MDM are explored and PSL and CL concepts are integrated to form new MDMs. Chorus (2010) and Bekhor et al (2012) introduce a Random Regret Minimization (RRM) model which assumes that individuals minimise anticipated regret, rather than maximize expected utility, when choosing routes, and Prato (2014) develops a Path Size RRM model and a Path Size Correction RRM model.

To summarise, the greatest hindrance for the GEV structures, Mixed Logit models, and alternative RUMs is the considerable computational cost required to solve the route choice probabilities for large-scale networks, while another issue that has been noted for some of the GEV structures and Mixed Logit models is the difficulty in obtaining reasonable estimates for parameters. MNL-modification models are a useful and practical approach to approximating the correlation; more complex models can capture the correlation more accurately, but due to the comparatively low computational cost and the relative ease in obtaining reasonable estimates for parameters, they are the most commonly used models in practice, and are the focus of this paper, in particular: Path Size Logit models.

One of the main issues for most MNL-modification models is that results are highly sensitive to the inclusion and exclusion of routes from the choice set. The CL model proposes that the correction terms are based upon commonality factors that measure the similarity of routes, and penalises the utilities accordingly. In contrast, the PSL and PSCL models propose that the correction terms are based upon path size terms that measure route distinctiveness: a route is penalised based on the number of other routes sharing its links, and the costs of those shared links. The inclusion of a route to a choice set can thus have a substantial effect upon the choice probability of any route that shares its links, as the correction terms adjust the route utilities attempting to capture the correlation.

In real-life applications where the size of the network is often large, route choice is rarely performed upon the full choice sets of routes, and choice sets are either pre-generated or a column generation approach is employed with implementation. This is often because it is computationally infeasible to both/either enumerate the full choice sets of routes and/or perform route choice upon the full choice sets of routes. Furthermore, typical road networks contain many very long routes that should be considered unrealistic and excluded from route choice. The key issue for the PSL model is that the choice probabilities are extremely sensitive to the utilised choice sets since all routes contribute equally to path size terms, and as such, results are extremely sensitive to the choice set generation / column generation method adopted. Moreover, it is crucial for the PSL model that the choice sets contain realistic alternatives only, as the inclusion of a single unrealistic alternative can have a considerable and negative effect on the choice probabilities of the realistic routes.

Thus, since it is difficult to obtain choice sets of realistic routes with absolute certainty for PSL to be suitable, a pragmatic approach is to utilise a weighted path size contribution technique along with choice set generation to attempt to reduce the impact any present unrealistic routes may have on the choice probabilities of realistic routes. Weighted path size contribution techniques weight the contribution of routes to path size terms with a path size contribution factor, i.e. so that the contribution of route k to the path size term of route i is reduced for unrealistic routes.

The GPSL model proposes that the path size contribution factor is based upon ratios of travel cost between routes, and hence routes with excessively large travel costs have a diminished impact upon the correction terms of routes with small travel costs, and consequently the choice probabilities of those routes. How the GPSL path size contribution factor

is formulated, however, means that: a) the model is not always internally consistent with how it assesses routes to be (un)realistic, and b) an additional scaling parameter is required to scale the path size contributions which makes parameter estimation more difficult. The main contribution of the present paper is thus the formulation of an Adaptive Path Size Logit (APSL) model where a new path size contribution factor is proposed so that the model is always internally consistent with how it assesses routes to be unrealistic, and so that no additional parameters are required for estimation.

For Path Size Logit models, the probability that a route is chosen (i.e. how feasible of an alternative the route is perceived to be) is a trade-off between its relative attractiveness due to travel cost and its relative attractiveness due to distinctiveness, and two independent scaling parameters (Logit parameter and path size parameter) affect the relativeness of how attractive each component is. The GPSL path size contribution factors however, are inconsistent with this in two ways: a) they assess the feasibility of a route according to its relative attractiveness due to travel cost only; and, b) they scale the relativeness of how attractive routes are due to travel cost with an additional independent path size contribution parameter, which is not necessarily proportional to the Logit parameter that scales travel cost within the probability relation.

The APSL model proposes that the path size contribution factors are based upon ratios of choice probability between routes, thus ensuring that routes defined as unrealistic by the path size terms, are exactly those with very low choice probabilities. The APSL route choice probability relation is an implicit function involving the choice probabilities, and solutions to the model are solutions to the fixed-point problem. Also, by defining the path size contribution factor as the ratio of choice probabilities, the scaling of the path size contributions is controlled implicitly through the scaling of the route choice probabilities (i.e. with the Logit parameter and path size parameter), and hence there is no additional path size contribution parameter for estimation.

Fig. 1A displays a network with a single OD movement and three routes. As x varies, the travel cost of each route stays constant (though different from one another), meaning that the GPSL path size terms always assess the feasibility of each route as being the same. However, as x is varied between 0 and 1, the correlation between Route 1 and Route 2 varies, thus altering the choice probabilities and how feasible of an alternative each route is perceived to be. A key point is that the behaviour of the choice probabilities as x is varied is highly dependent upon the values of the scaling parameters; different ranges for the Logit / path size parameters imply different theoretical route choice behaviours consequently altering how feasible each route is deemed to be. This makes it difficult to state what behaviours we should expect to happen as x is varied, without knowing the parameter values we wish to set, to model specific behaviours. Due to its internal consistency, the APSL model is adaptable to whichever values are set for the scaling parameters, and the path size contribution factors will assess the feasibility of routes by how relatively attractive they are due to travel cost and distinctiveness as the scaling parameters dictate.

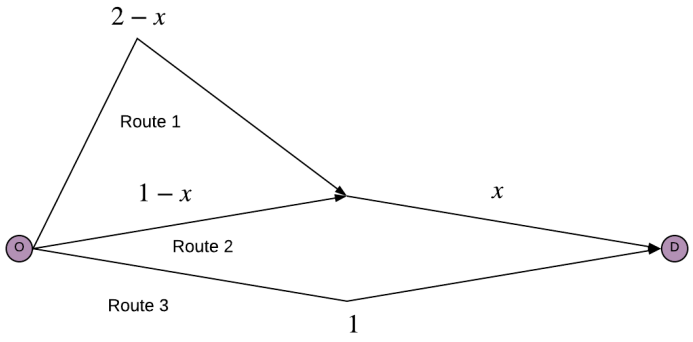


Fig. 1. Example network to demonstrate the inconsistency of the GPSL model.

The structure of the paper is as follows. In Section 2 we introduce some basic network notation as well as the definitions of the MNL, PSL, and GPSL models, and several numerical experiments on small-scale networks to demonstrate some of the key issues with PSL and GPSL, and the potential negative implications of an internally inconsistent PSL model. In Section 3 we detail the new APSL model, give results from several numerical experiments that demonstrate the key properties of the APSL model, and detail a solution method. Section 4 addresses existence and uniqueness of APSL solutions. In Section 5 we investigate estimating the APSL model. To show that the parameters of the APSL model can be estimated we first propose a Maximum Likelihood Estimation procedure for estimating APSL with tracked route observation data, then investigate this procedure in a simulation study on the Sioux Falls network where we show that it is generally possible to reproduce assumed true parameters. Then, in a real-life case study, we estimate the APSL model using real tracked route GPS data on a large-scale network. Section 6 concludes the paper.

2 Notation, Definitions, and Demonstrations of Key Issues with Existing Path Size Logit Models

2.1 Basic Network Notation

The model developed in this paper is applicable to general networks with multiple OD movements and flow-dependent link costs. However, without compromising the model derivation, we simplify notation by considering a single OD movement with fixed link costs. The network consists of link set A. For the OD movement, R is the choice set of all simple routes (without cycles), having size $N = |R|$. $A_i \subseteq A$ is the set of links belonging to route $i \in R$, and $\delta_{a,i} = \begin{cases} 1 & if\ a \in A_i \\ 0 & otherwise \end{cases}$. Each link $a \in A$ has a fixed generalised travel cost t_a, and supposing that the travel cost for a route can be attained through summing up the total cost of its links, then the generalised travel cost for route $i \in R$, c_i, can be computed as follows: $c_i = \sum_{a \in A_i} t_a$.

The route choice probability for route $i \in R$ is P_i, where $\boldsymbol{P} = (P_1, P_2, \dots, P_N)$ is the vector of route choice probabilities, and D is the set of all possible route choice probability vectors:

$$D = \left\{ \boldsymbol{P} \in \mathbb{R}^N_{\geq 0} : 0 \leq P_i \leq 1, \forall i \in R, \sum_{j=1}^{N} P_j = 1 \right\}.$$

And, $D^{>0} \subset D$ is the subset of all possible route choice probability vectors where no route has zero choice probability:

$$D^{>0} = \left\{ \boldsymbol{P} \in \mathbb{R}^N_{>0} : 0 < P_i < 1, \forall i \in R, \sum_{j=1}^{N} P_j = 1 \right\}.$$

2.2 Multinomial Logit

The Multinomial Logit (MNL) choice model is formulated as follows. The deterministic utility of alternative $i \in R$ is V_i, and the random utility of alternative $i \in R$ is U_i such that $U_i = V_i + \varepsilon_i$, where the ε_i terms are the individually and identically distributed random variable error terms. Assuming individuals seek the alternative with highest utility, the probability that an individual selects alternative $i \in R$ is:

$$P_i = \Pr\left(U_i \geq U_j, \forall j \in R, j \neq i\right) = \Pr\left(V_i + \varepsilon_i \geq V_j + \varepsilon_j, \forall j \in R, j \neq i\right).$$

The defining characteristic of Logit models is that the random variable error terms assume a Gumbel distribution. Consequently:

$$P_i(\boldsymbol{V}) = \frac{e^{V_i}}{\sum_{j \in R} e^{V_j}}$$

where $\boldsymbol{V}$ is the vector of deterministic utilities.

The MNL model in the context of route choice states that the deterministic utility of route $i \in R$ is given by $V_i = -\theta c_i$, where $\theta > 0$ is the Logit scaling parameter, and thus:

$$P_i = \frac{e^{-\theta c_i}}{\sum_{j \in R} e^{-\theta c_j}} = \frac{1}{\sum_{j \in R} e^{-\theta (c_j - c_i)}}. \quad (1)$$

The MNL model assumes the route utilities are independent from one another, however routes with overlapping links share unobserved attributes, and the assumption that the random error terms are all independently and identically distributed is no longer valid. The famous example for this is the 'loop hole' network (also known as the red-bus/blue-bus network) presented in Cascetta et al (1996).

2.3 Path Size Logit Models

Path Size Logit models include correction terms to penalise routes that share links with other routes, so that the deterministic utility of route $i \in R$ is $V_i = -\theta c_i + \kappa_i$, where $\kappa_i \leq 0$ is the correction term for route $i \in R$. The probability that a driver chooses route $i \in R$ is therefore:

$$P_i = \frac{e^{-\theta c_i + \kappa_i}}{\sum_{j \in R} e^{-\theta c_j + \kappa_j}}.$$

Path Size Logit models adopt the form $\kappa_i = \beta \ln(\gamma_i)$, where $\beta \geq 0$ is the path size scaling parameter, and $\gamma_i \in (0,1]$ is the path size term for route $i \in R$. A distinct route with no shared links has path size term equal to 1, resulting in no penalisation. Less distinct routes have smaller path size terms and incur greater penalisation. The probability that a driver chooses route $i \in R$ is:

$$P_i = \frac{e^{-\theta c_i + \beta \ln(\gamma_i)}}{\sum_{j \in R} e^{-\theta c_j + \beta \ln(\gamma_j)}} = \frac{(\gamma_i)^{\beta} e^{-\theta c_i}}{\sum_{j \in R} (\gamma_j)^{\beta} e^{-\theta c_j}} = \frac{1}{\sum_{j \in R} \left(\frac{\gamma_j}{\gamma_i}\right)^{\beta} e^{-\theta(c_j - c_i)}}. \quad (2)$$

2.3.1 Path Size Logit

The Path Size Logit (PSL) model was first proposed by Ben-Akiva & Ramming (1998), and states that the PSL path size term for route $i \in R$, γ_i^{PS}, is defined as follows:

$$\gamma_i^{PS} = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R} \delta_{a,k}}. \quad (3)$$

To dissect the PSL path size term for route $i \in R$ defined in (3): each link a in route i is penalised (in terms of decreasing the path size term and hence the utility of the route) according to the number of routes in the choice set that also use that link $\left(\sum_{k \in R} \delta_{a,k}\right)$, and the significance of the penalisation is weighted according to how prominent link a is in route i, i.e. the cost of route a in relation to the total cost of route i $\left(\frac{t_a}{c_i}\right)$.

PSL Key Issue: *Unrealistic routes negatively impact the choice probabilities of realistic routes when links are shared.*

The key issue with the PSL model is that all routes contribute equally to path size terms (i.e. the path size contribution factors are simply all 1), and hence the choice probabilities of realistic routes are affected by link sharing with unrealistic routes. To demonstrate this, consider example network 1 in Fig. 2A where there are 3 routes: Routes 2 & 3 have travel cost 1 and Route 1 has travel cost $1 + \upsilon$, Routes 1 & 2 are correlated while Route 3 is distinct. Fig. 2B displays the example network 1 PSL choice probabilities as υ is increased from 0.5 to 3, $\theta = \beta = 1$. For $\upsilon = 0.5$, Routes 1 & 2 have the same unshared travel cost and are thus considered equally attractive. As υ is increased, Route 1 increases in travel cost and decreases in choice probability. As Route 1 becomes an unrealistic alternative, the choice probability of Route 2 should not be penalised for overlapping with Route 1. The PSL path size terms dictate though that Route 1 contributes equally to the path size term of Route 2 for all υ, and hence the choice probability of Route 2 is always significantly penalised. As υ is increased and the choice probability of Route 1 approaches zero, the contribution of Route 1 to the path size term of Route 2 should decrease, and the choice probability of Route 2 should converge to the choice probability of Route 3.

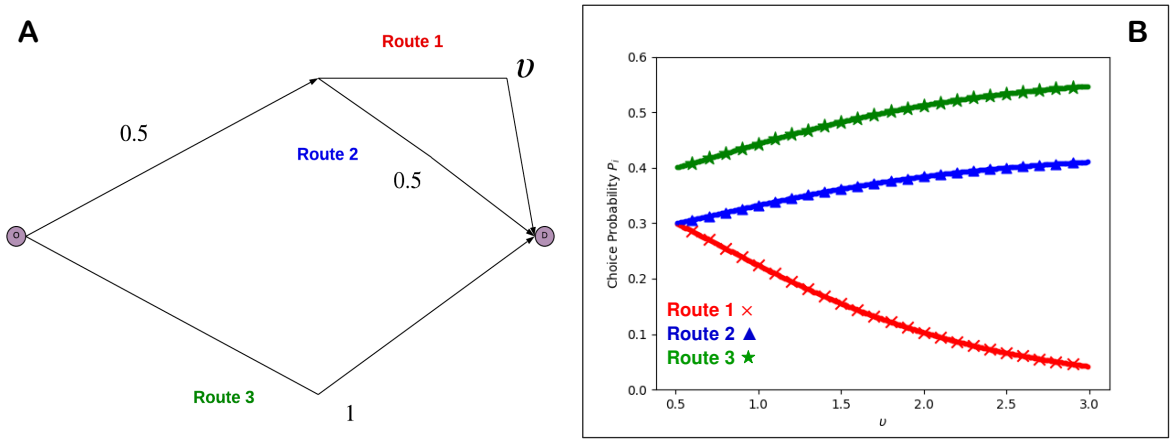


Fig. 2. A: Example network 1. B: Example network 1: PSL route choice probabilities for increasing υ; $\theta = \beta = 1$.

2.3.2 Generalised Path Size Logit

Ben-Akiva & Bierlaire (1999) formulate an alternative PSL model (PSL′) that attempts to reduce the contributions of excessively expensive routes to the path size terms of more realistic routes in the choice set. The PSL′ model states that the PSL′ path size term for route $i \in R$, $\gamma_i^{PS'}$, is defined as follows:

$$\gamma_i^{PS'} = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R} \left(\frac{\min(c_l : l \in R)}{c_k}\right) \delta_{a,k}}, \quad (4)$$

As (4) shows, the contribution of route k to path size terms is weighted according to the ratio of route k and the cheapest route in the choice set $\left(\frac{\min(c_l : l \in R)}{c_k}\right)$, and hence contributions of high costing routes compared to the cheapest alternative are reduced.

As Ramming (2002) describes, however, when a route is completely distinct its path size term is not always equal to 1 which results in an undesired penalisation upon the utility of that route. To combat this, Ramming (2002) proposes the Generalised Path Size Logit (GPSL) model. The GPSL model states that the GPSL path size term for route $i \in R$, γ_i^{GPS}, is defined as follows:

$$\gamma_i^{GPS} = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R} \left(\frac{c_i}{c_k}\right)^{\lambda} \delta_{a,k}}, \tag{5}$$

where $\lambda \geq 0$, noting that the GPSL model is equivalent to the PSL model when $\lambda = 0$. In (5), the contribution of route k to the path size term of route i (the path size contribution factor) is weighted according to the cost ratio between the routes, $\left(\left(\frac{c_i}{c_k}\right)^{\lambda}\right)$, and hence the contributions of high costing routes to the path size terms of low costing routes is reduced. $\lambda \geq 0$ is the path size contribution scaling parameter to be estimated.

Fig. 3 displays the example network 1 GPSL choice probabilities as υ is increased from 0.5 to 3, $\theta = \beta = 1$, $\lambda = 3$. As Fig. 3 shows, as υ is increased and the travel cost of Route 1 increases, the contribution of Route 1 to the path size term of Route 2 decreases, and consequently the choice probability of Route 2 converges to the choice probability of Route 3.

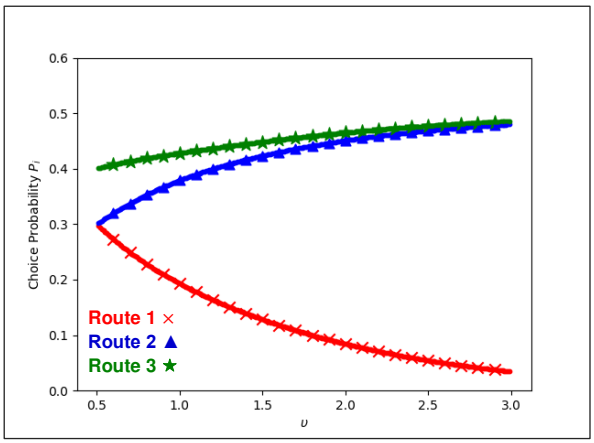


Fig. 3. Example network 1: GPSL route choice probabilities for increasing υ; $\theta = \beta = 1$, $\lambda = 3$.

GPSL Key Issue 1: *For large λ, GPSL path size terms are highly sensitive to small differences in route travel cost.*

It is mentioned numerous times in the literature that the GPSL model can be problematic for large λ values, especially when overlapping routes only have marginally different travel costs (Ramming, 2002; Frejinger & Bierlaire, 2007; Hoogendoorn-Lanser, 2005). Hoogendoorn-Lanser et al (2005) describe how λ should be set to 0 when overlapping routes have more-or-less equal travel costs, as the overlap between those alternatives should not affect their choice probabilities differently. However, when overlapping routes have very different travel costs λ should not be set to 0, as the effects that routes with high travel costs have on the path size terms of routes with low travel costs should be dampened. Example network 2 in Fig. 4A shows a network where both cases exist: Routes 1 & 2 are overlapping routes with more-or-less equal travel costs ($c_1 = 2.01$, $c_2 = 2$), and Routes 3 & 4 are overlapping routes with very different travel costs ($c_3 = 2$, $c_4 = 6$). Fig. 4B shows the example network 2 GPSL choice probabilities as λ is increased from 0 to 400, $\theta = \beta = 1$. When $\lambda = 0$ (i.e. GSPL is equivalent to PSL), Routes 1, 2 & 3 have approximately equal choice probabilities as they all have similar travel costs and all share approximately half of their journey with one other route. Route 4 induces a penalty on Route 3, but this should be less than the path size penalties Routes 1 & 2 impose on each other, and thus P_3 should be greater than P_1 and P_2, which should be approximately equal, for these values of θ and β. Although this is the case when $\lambda \cong 10$, increasing λ amplifies the difference in costs between Routes 1 & 2 so that P_1 and P_2 diverge, which is not desired.

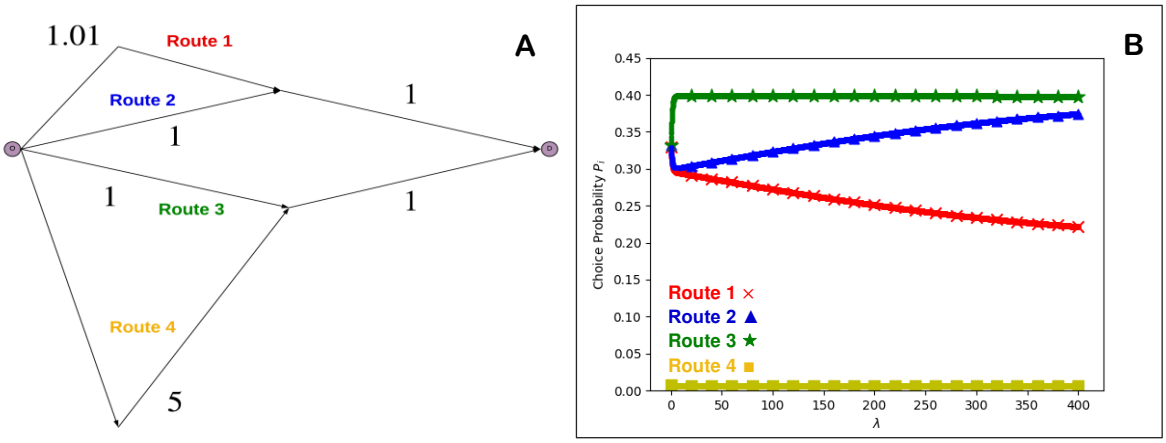


Fig. 4. A: Example network 2. B: Example network 2: GPSL route choice probabilities for increasing λ; $\theta = \beta = 1$.

As (2) shows, how feasible route i is perceived to be by drivers (i.e. its choice probability) is a trade-off between its relative attractiveness due to travel cost $e^{-\theta(c_j - c_i)}$ and its relative attractiveness due to distinctiveness $\left(\frac{\gamma_j}{\gamma_i}\right)^{\beta}$, where the θ parameter scales relative attractiveness due to travel cost and β scales relative attractiveness due to distinctiveness. The GPSL path size contribution factors in (4), however, assess the feasibility of route i according to its relative attractiveness due to travel cost only $\left(\frac{c_i}{c_k}\right)^{\lambda}$, where an additional parameter λ scales relative attractiveness due to travel cost. The internal inconsistency issues of the GPSL model between the probability relation and path size terms are thus twofold: a) there is an inconsistent assessment of the feasibility of routes; and, b) there are inconsistent parameters that scale relative attractiveness due to travel cost. We demonstrate both issues below.

GPSL Key Issue 2: *Internally inconsistent assessment of the feasibility of routes.*

Fig. 5A shows a network with 6 routes: Routes 1-5 are highly correlated with each other with fixed travel cost 1.2, and Route 6 has a fixed cost of 2 and is either a distinct route when $\eta = 1$ or correlated with Routes 1-5 when $\eta < 1$. As η is decreased, the correlation between Route 6 and Routes 1-5 increases, but the travel costs don't change. Supposing that $\theta = \beta = 1$, then one might expect Route 6 to have a choice probability greater than each of Routes 1-5 when $\eta = 1$ due to being distinct, and a choice probability smaller than each of Routes 1-5 when $\eta = 0$ due to having the larger detour. There should thus be a point $\eta_{eq} \in (0,1)$ where all routes have equal choice probability. At this point, each route is considered equally attractive and all routes should contribute equally to path size terms. By the definition of the PSL model all routes always contribute equally to path size terms; Fig. 5B shows the example network 3 PSL choice probabilities for increasing η, $\theta = \beta = 1$, and $\eta_{eq} = 0.27$ is the point where all routes have equal choice probabilities. For $\eta \neq \eta_{eq}$ however, it is not necessarily required for routes to contribute equally to path size terms, for example when $\eta = 0$ the contribution of Route 6 to Routes 1-5 may wish to be reduced, and the PSL model is incapable of this. Fig. 5C & Fig. 5D show the example network 3 GPSL choice probabilities for increasing η, $\theta = \beta = 1$, for $\lambda = 1$ and $\lambda = 10$, respectively. The GPSL model proposes that the contribution of Route 6 to the path size terms of Routes 1-5 is a constant $\left(\frac{1.2}{2}\right)^{\lambda}$ for all η, and thus the points where all routes have equal choice probabilities are $\eta_{eq} = 0.37$ and $\eta_{eq} = 0.48$ for $\lambda = 1$ and $\lambda = 10$, respectively, which are greater than 0.27, and larger values for λ moves η_{eq} further away from 0.27.

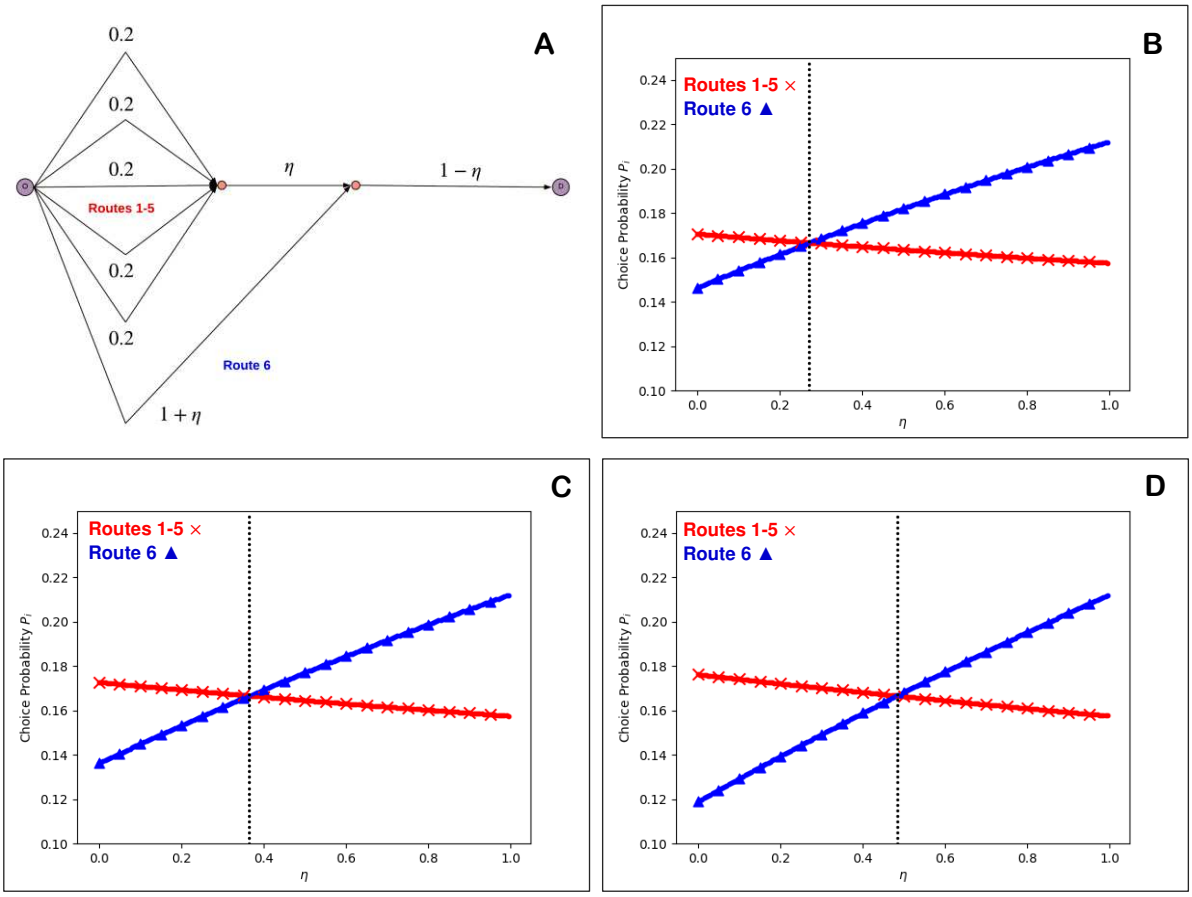


Fig. 5. A: Example network 3. Example network 3: Route choice probabilities for increasing η; $\theta = \beta = 1$ – B: PSL, $(\eta_{eq} = 0.27)$. C: GPSL, $\lambda = 1$, $(\eta_{eq} = 0.37)$. D: GPSL, $\lambda = 10$, $(\eta_{eq} = 0.48)$.

GPSL Key Issue 3: *Internally inconsistent scaling parameters.*

Fig. 6A displays a network with 4 routes: Routes 1 & 4 have travel cost 3 and Routes 2 & 3 have travel cost 1, Routes 1 & 2 are distinct and Routes 3 & 4 are correlated, with Route 4 being more distinct than Route 3. Different ranges for the θ and β parameters have different implications for how relatively attractive the routes are due to travel cost and distinctiveness. Fig. 6B shows the example network 4 GPSL choice probabilities for increasing θ, $\beta = 1$, $\lambda = 4$. The GPSL model diminishes Route 4's contribution to the path size term of Route 3 to $\left(\frac{1}{3}\right)^4$, and thus Routes 2 & 3 have near identical choice probabilities for all θ. For low θ however, the sensitivity to the differences in travel cost is dampened within the probability relation, yet Route 4's path size term contribution to Route 3 accentuates the difference in cost, and the GPSL model is thus inconsistent.

To try and overcome this inconsistency issue, one must attempt to represent λ proportional to θ. Because θ scales travel cost difference, and λ scales travel cost ratios, it is difficult to know how λ should relate to θ, e.g. $\lambda = \theta^5$? A potential solution could be to adjust the GPSL path size contribution factor to resemble the relative attractiveness due to travel cost component within the Path Size Logit probability relation in (2) (and the MNL probability relation in (1)), thus formulating an alternative Generalised Path Size Logit (GPSL′) model, where the GPSL′ path size term for route $i \in R$, $\gamma_i^{GPS'}$, is defined as follows:

$$\gamma_i^{GPS'} = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R} e^{-\lambda(c_k - c_i)} \delta_{a,k}}. \tag{6}$$

By setting $\lambda = \theta$, the relative attractiveness due to travel cost components within the GPSL′ model match exactly. Fig. 6C shows the example network 4 GPSL′$_{(\lambda=\theta)}$ choice probabilities for increasing θ, $\beta = 1$. For low θ, the sensitivity to the difference in cost between Routes 3 & 4 is dampened within the probability relation *and* within the path size contribution factor, and Route 3 has the lowest choice probability due to being the least distinct. As θ is increased, the sensitivity to the difference in cost between Routes 3 & 4 is accentuated within the path size contribution factor and Route 4's contribution to Route 3 decreases, and the choice probability of Route 3 converges to the choice probability of Route 2.

The GPSL′$_{(\lambda=\theta)}$ model only partially improves the internal consistency of the GPSL model though. Fig. 6D shows the example network 4 GPSL′$_{(\lambda=\theta)}$choice probabilities for increasing θ, $\beta = 2$. An increase in β increases the sensitivity to distinctiveness within the probability relation; attractiveness due to distinctiveness is not considered within the GPSL & GPSL′ path size contribution factors however, and hence as Route 3 becomes unattractive for low θ due to being the least distinct, its path size contribution to Route 4 is not reduced, though approaching being considered unrealistic as the parameters dictate.

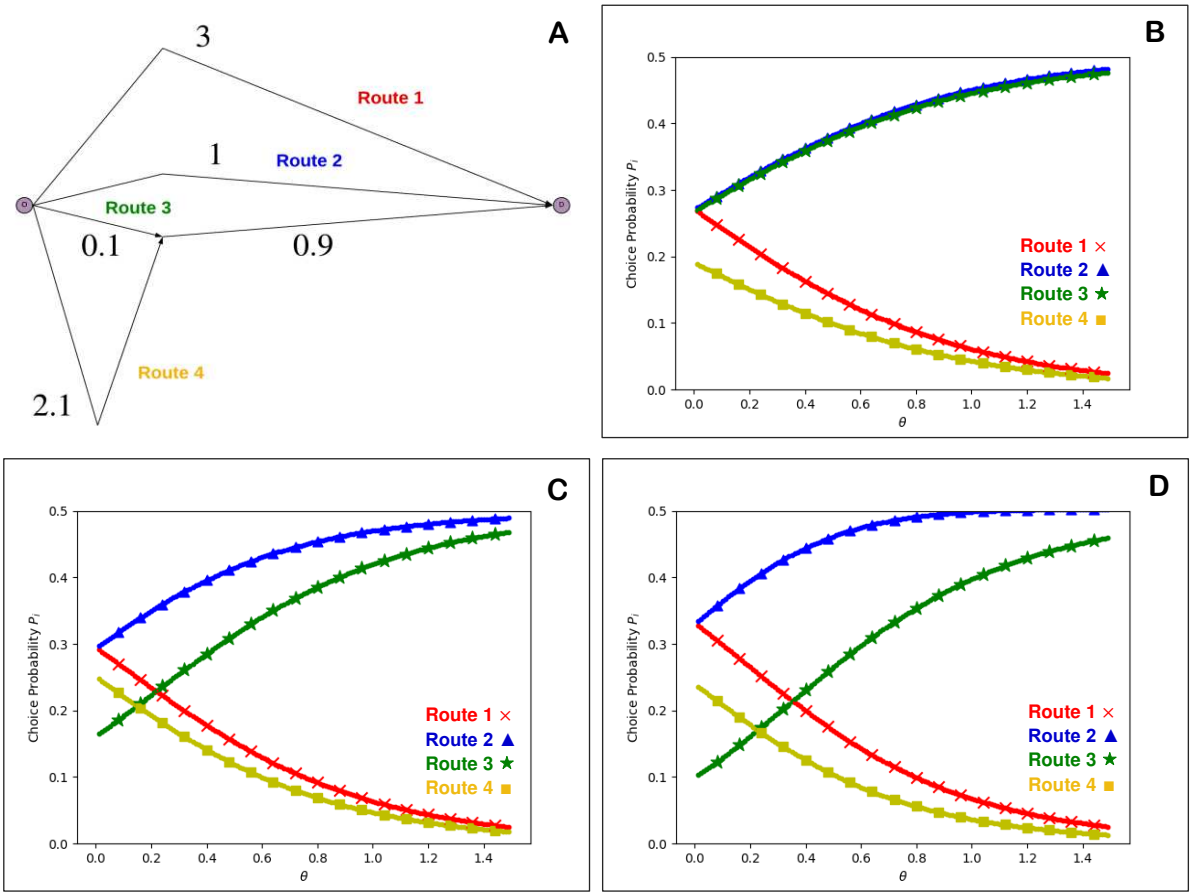


Fig. 6. A: Example network 4. Example network 4: Route choice probabilities for increasing θ – B: GPSL, $\beta = 1, \lambda = 4$. C: GPSL′, $\lambda = \theta, \beta = 1$. D: GPSL′, $\lambda = \theta, \beta = 2$.

3 The Adaptive Path Size Logit Model

In the PSL model, all routes contribute equally to path size terms so that unrealistic routes negatively impact the choice probabilities of realistic routes when links are shared. The GPSL model attempts to overcome this issue by including a path size contribution factor based upon travel cost ratios, but has issues with internal inconsistency. The alternative GPSL (GPSL′$_{(\lambda=\theta)}$) model partially addresses this inconsistency but does not take into account the relative attractiveness of routes due to distinctiveness. We thus propose in this section a fully internally consistent PSL model where all components assess the feasibility of routes according to its relative attractiveness due to travel cost and distinctiveness. Formulation of the APSL model was complicated by the desire to establish existence and uniqueness of solutions. First, we provide a simpler formulation of the APSL model, then we detail the final more complicated definition, which has been constructed solely so that solutions exist and can be unique (proven in Section 4).

3.1 Preliminary Definition of APSL

Definition

The preliminary definition of APSL (APSL$_0$) is defined as follows. The APSL$_0$ choice probabilities, $\boldsymbol{P}^*$, are a solution to the fixed-point problem $\boldsymbol{P} = \boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)$, where:

$$g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right) = \frac{e^{-\theta c_i+\beta \ln\left(\gamma_i^{APS}(\boldsymbol{P})\right)}}{\sum_{j\in R} e^{-\theta c_j+\beta \ln\left(\gamma_j^{APS}(\boldsymbol{P})\right)}} = \frac{\left(\gamma_i^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_i}}{\sum_{j\in R}\left(\gamma_j^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_j}} = \frac{1}{\sum_{j\in R}\left(\frac{\gamma_j^{APS}(\boldsymbol{P})}{\gamma_i^{APS}(\boldsymbol{P})}\right)^{\beta} e^{-\theta(c_j-c_i)}}, \tag{7}$$

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum\nolimits_{a\in A_i} \frac{t_a}{c_i}\frac{P_i}{\sum_{k\in R} P_k\delta_{a,k}} = \sum\nolimits_{a\in A_i} \frac{t_a}{c_i}\frac{1}{\sum_{k\in R}\left(\frac{P_k}{P_i}\right)\delta_{a,k}}, \qquad \forall i \in R, \tag{8}$$

$$\forall \boldsymbol{P} \in D^{>0}, \qquad \theta > 0, \qquad \beta \geq 0.$$

$\gamma_i^{APS}(\boldsymbol{P})$ in (8) is the APSL path size term function for route $i \in R$ that is a function involving the route choice probabilities. $g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)$ in (7) is the APSL route choice probability function for route $i \in R$ which is a function involving the path size term functions and hence also the choice probabilities of routes. The choice probability relation for route $i \in R$ is given by $P_i = g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)$, which is an implicit equation involving choice probabilities, and hence the APSL$_0$ route choice probabilities, $\boldsymbol{P}^*$, are a solution such that $P_i^* = g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P}^*)\right), \forall i \in R$.

Property 1: For an APSL$_0$ route choice probability solution vector $\boldsymbol{P}^*$, $\gamma_i^{APS}(\boldsymbol{P}^*)$ is the APSL path size term for route $i \in R$, $\kappa_i = \beta \ln\left(\gamma_i^{APS}(\boldsymbol{P}^*)\right)$ is the correction term, and the deterministic utility is given by:

$$V_i = -\theta c_i + \beta \ln\left(\gamma_i^{APS}(\boldsymbol{P}^*)\right). \tag{9}$$

If the random utility for route $i \in R$ is $U_i = -\theta c_i + \beta \ln\left(\gamma_i^{APS}(\boldsymbol{P}^*)\right) + \varepsilon_i$, and if the random variable error terms, ε_i, are i.i.d Gumbel, then the probability relation in (7) is obtained.

As (8) shows, for a choice probability solution $\boldsymbol{P}^*$, the contribution of route k to the APSL path size term of route i is weighted according to the ratio of choice probabilities between the routes $\left(\frac{P_k^*}{P_i^*}\right)$, and hence unrealistic route alternatives with very low choice probabilities have a diminished contribution to the path size terms of realistic routes with relatively large choice probabilities. The choice probability ratio path size contribution factor can also be formulated as follows:

$$\frac{P_k^*}{P_i^*} = \frac{g_k\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P}^*)\right)}{g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P}^*)\right)} = \left(\frac{\gamma_k^{APS}(\boldsymbol{P}^*)}{\gamma_i^{APS}(\boldsymbol{P}^*)}\right)^{\beta} e^{-\theta(c_k-c_i)}. \tag{10}$$

So, it is clear to see that the path size contribution factor in (10) matches the probability relation in (7), where both consider how attractive a route is by measuring its relative attractiveness due to travel cost and distinctiveness, and hence there is some clear consistency within the model's framework. Furthermore, not only do the path size contribution factors become more consistent with the eventual route choice probabilities (i.e. they both define a route as being unrealistic if it has a relatively unattractive combination of travel cost and distinctiveness) the model does not require the estimation of any additional parameters. Whereas the scaling of the path size contributions in the GPSL model depends upon an independent parameter λ, the scaling of the path size contributions in the APSL$_0$ model depends implicitly on the scaling of the choice probability relation with the θ and β parameters, and one cannot independently adjust the scaling within the path size contribution factors without scaling the choice probability relation as well.

In Random Utility Theory (RUT), RUMs are derived based on the deterministic utility function and random error term. As (9) shows, however, the deterministic utility function for the APSL$_0$ model is not in fact deterministic since it is dependent upon the route choice probabilities. APSL$_0$ is thus not a member of the RUM family, though it is derived using RUT. This is analogous to the way that Stochastic User Equilibrium (SUE) is a consistency condition derived using RUT but an SUE model is not a member of the RUM family. The key difference can be understood in considering a policy test: having solved the APSL$_0$ fixed-point problem for the 'before' case, the path size terms are not fixed. The 'after' case would require APSL$_0$ to be re-solved and the path size terms would be updated. If instead one fixed the APSL path size terms from the 'before' case, then the 'after' case would be a regular Path Size Logit model and hence a RUM. Making this explicit, imagine examining the impact on route choice of a new road added to the network. For pre-existing routes not overlapping with the new routes generated, the path size terms of the regular Path Size Logit models would remain the same, and hence so does the attractiveness of those routes. For the APSL$_0$ model, however, the fixed-point probability system must be re-solved, and the path size terms for *all* pre-existing routes may be adjusted to account for the updated attractiveness of the routes, and thus to alter the path size contributions.

The Issue

Standard proofs for existence and uniqueness of fixed-point solutions require the domain of the fixed-point function (in this case $\boldsymbol{g}$) to be a compact convex set. The issue with the APSL$_0$ model as defined in (7) and (8) is that the domain of the fixed-point function $\boldsymbol{g}$, $D^{>0}$, is open and bounded (not compact) as the function is not always defined for zero choice

probabilities where $\frac{0}{0}$ can occur in the path size terms. Altering the definition of the path size terms so that the domain of the fixed-point function $\boldsymbol{g}$ is the closed and bounded set D, where routes can have zero choice probabilities, is however problematic as issues arise with ensuring that $\boldsymbol{\gamma}^{APS}(\boldsymbol{P})$ remains a continuous function (see Appendix A for a demonstration).

3.2 Proposed APSL Definition

While the preliminary definition of the APSL model defined in the previous subsection is the model we originally aimed to propose, as discussed, we could not prove that solutions were guaranteed to exist nor be unique according to standard proofs, and hence we provide here an altered model for use where solutions are guaranteed to exist and where conditions for uniqueness can be defined.

Definition

The APSL choice probabilities, $\boldsymbol{P}^*$, (for a choice set of size N) are a solution to the fixed-point problem $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$, where:

$$G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right) = \tau + (1 - N\tau) \cdot g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right), \tag{11}$$

$$g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right) = \frac{e^{-\theta c_i + \beta \ln\left(\gamma_i^{APS}(\boldsymbol{P})\right)}}{\sum_{j \in R} e^{-\theta c_j + \beta \ln\left(\gamma_j^{APS}(\boldsymbol{P})\right)}} = \frac{\left(\gamma_i^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_i}}{\sum_{j \in R}\left(\gamma_j^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_j}}, \tag{12}$$

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{P_i}{\sum_{k \in R} P_k \delta_{a,k}} = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R}\left(\frac{P_k}{P_i}\right)\delta_{a,k}}, \tag{13}$$

$$\forall i \in R, \qquad \forall \boldsymbol{P} \in D^{(\tau)}, \qquad \theta > 0, \qquad \beta \geq 0, \qquad 0 < \tau \leq \frac{1}{N},$$

$$D^{(\tau)} = \left\{\boldsymbol{P} \in \mathbb{R}^N_{>0} : \tau \leq P_i \leq (1 - (N-1)\tau), \forall i \in R, \sum_{j=1}^{N} P_j = 1\right\}.$$

(11) and (12) are equivalent to (7) and (8) for the preliminary definition: $\gamma_i^{APS}(\boldsymbol{P})$ in (13) is the APSL path size term function for route $i \in R$ that is a function involving the route choice probabilities, and $g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)$ in (12) is the unadjusted choice probability function for route $i \in R$ which is a function involving the path size term functions and hence also the choice probabilities of routes. The choice probability relation for route $i \in R$ is given by $P_i = G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$, which is an implicit equation involving choice probabilities, and hence the APSL route choice probabilities, $\boldsymbol{P}^*$, are a solution such that $P_i^* = G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P}^*)\right)\right)$, $\forall i \in R$. The key difference between this final model here and the preliminary definition is the adjustment function G_i. $G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ in (11) is the APSL choice probability adjustment function for route $i \in R$ which adjusts the choice probability function g_i for reasons given below.

Motivation

G_i is a fixed-point function, and its construction was motivated by some desired behaviours, as well as some required properties for proving existence and uniqueness:

1. G_i must map into itself.
2. G_i must be continuous for all $\boldsymbol{P}$.
3. G_i must be continuously differentiable with respect to $\boldsymbol{P}$ for all $\boldsymbol{P}$.
4. The domain of G_i must be closed and bounded.
5. The domain of G_i must not allow for zero choice probabilities.
6. G_i should be able to approximate g_i.
7. The domain of G_i should allow for choice probabilities to be approximately close to zero.

Desired Properties (DP) 1-4 are required for existence and uniqueness proofs. DP 5 is required since the path size term function $\gamma_i^{APS}(\boldsymbol{P})$ in (13) and thus $G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ in (11) can be undefined for zero choice probabilities where $\frac{0}{0}$ can occur. DP 6 is desired as it is not our intention for the choice probabilities acquired from this final APSL model to be different to the choice probabilities from the preliminary definition (where one would exist), and we wish them to be as

close as possible. DP 7 is desired as we wish to be able to obtain choice probability solutions where there are unrealistic routes with extremely small choice probabilities.

So, the formulation of G_i in (11) and its domain $D^{(\tau)}$ have been constructed to satisfy DP 1-7. In Section 4 we prove that DP 1-3 are satisfied. The parameter τ is introduced, and the domain $D^{(\tau)}$ is such that $P_i \geq \tau, \forall i \in R$, and since the choice probabilities for all routes sum up to 1: $P_i \leq (1-(N-1)\tau), \forall i \in R$. DP 4 is thus satisfied as $D^{(\tau)}$ is closed and bounded. Moreover, τ is restricted to the range $0 < \tau \leq \frac{1}{N}$ and thus DP 5 is satisfied as zero choice probabilities are not in the domain. As $\tau \to 0$, $G_i \to g_i$ satisfying DP 6 and the lower bound for P_i in $D^{(\tau)}$ tends towards zero satisfying DP 7.

It is important to note that **the τ parameter is not a model parameter that requires estimating**, it is simply a mathematical construct that ensures DP 1-7 are satisfied. While the proposed APSL model provides the capability, it is not our intention for this final APSL model to purposefully compute different choice probabilities to those obtained from the preliminary definition for any given theoretical reason. In fact, we desire the choice probabilities to be as close as possible, and hence we advise that only small values of τ are used. For the rest of the paper, i.e. for the demonstrations and estimation work, we set $\tau = 10^{-16}$, unless stated otherwise. In Section 5.3.2.2 we briefly investigate the impact of the τ parameter upon parameter estimation.

3.3 Demonstrations of Key Properties

APSL Key Property 1: *Unrealistic routes have a diminished impact upon the choice probabilities of realistic routes when links are shared.*

Referring to the PSL Key Issue in Section 2.3.1, Fig. 7 displays the example network 1 APSL choice probabilities as υ is increased from 0.5 to 3, $\theta = \beta = 1$. As Fig. 7 shows, as the travel cost of Route 1 increases, so does the relative unattractiveness of Route 1, thus decreasing the choice probability for that route and its influence upon the correction term of Route 2. The choice probability of Route 2 thus converges to match the choice probability of Route 3 as υ is increased.

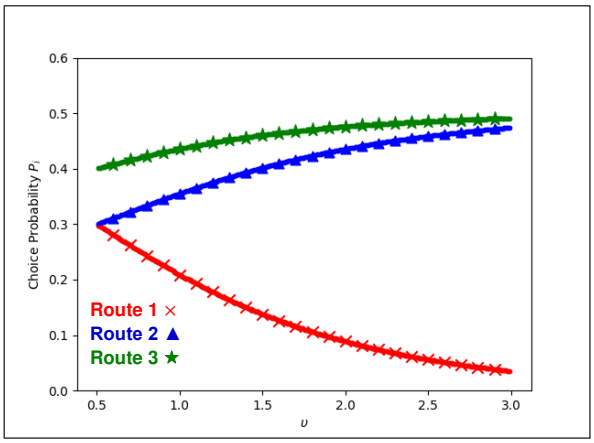


Fig. 7. Example network 1: APSL route choice probabilities for increasing υ; $\theta = \beta = 1$.

APSL Key Property 2: *APSL path size terms aren't highly sensitive to small differences in route travel cost.*

Referring to GPSL Key Issue 1 in Section 2.3.2, it was demonstrated how GPSL path size terms can be highly sensitive to small differences in route travel cost when λ values are large. It was also demonstrated how setting $\lambda = 0$ (i.e. equivalating GPSL and PSL) can also have adverse effects, as the routes with large travel costs negatively impact the choice probabilities of routes with low travel costs when links are shared. For example network 2 in Fig. 4A, a good compromise was found by setting $\lambda = 10$. APSL path size terms do not suffer from the same issues GPSL path size terms have when $\lambda = 0$ and when λ is large. Table 1 displays the example network 2 GPSL choice probabilities when $\lambda = 0$, $\lambda = 10$, and $\lambda = 400$, as well as the APSL choice probabilities, $\theta = \beta = 1$. As Table 1 shows, the APSL choice probabilities resemble the 'optimised' GPSL choice probabilities for $\lambda = 10$.

	PSL (= GPSL $\lambda = 0$)	GPSL ($\lambda = 10$)	GPSL ($\lambda = 400$)	APSL
P_1	0.329	0.294	0.222	0.297
P_2	0.332	0.301	0.374	0.301
P_3	0.332	0.399	0.398	0.397

P_4	0.007	0.006	0.006	0.006

Table 1. Example network 2: choice probabilities for different PSL models; $\theta = \beta = 1$.

APSL Key Property 3: *Internally consistent assessment of the feasibility of routes.*

Referring to GPSL Key Issue 2 in Section 2.3.2 where it was demonstrated how the GPSL model is not internally consistent with the assessment of the feasibility of routes, Fig. 8 shows the example network 3 APSL choice probabilities for increasing η, $\theta = \beta = 1$. At the point $\eta = \eta_{eq} = 0.27$, all routes have equal choice probabilities; the APSL model proposes that routes contribute to path size terms according to choice probability ratios and thus the path size contributions all cancel out at that point resulting in the PSL model, as desired.

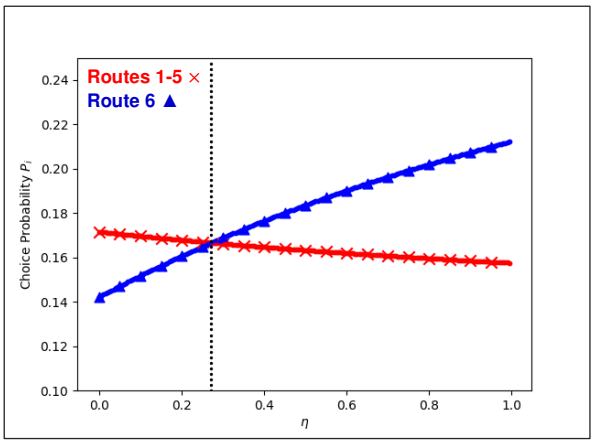


Fig. 8. Example network 3: APSL route choice probabilities for increasing η; $\theta = \beta = 1$.

APSL Key Property 4: *Internally consistent scaling parameters.*

Referring to GPSL Key Issue 3 in Section 2.3.2, it was demonstrated how the GPSL model has internally inconsistent scaling parameters, and how the GPSL$'_{(\lambda=\theta)}$ model partially improves the internal consistency of the GPSL model. Fig. 9A shows the example network 4 APSL choice probabilities for increasing θ, $\beta = 1$. As shown in (7) and (10), the APSL model, like the GPSL$'_{(\lambda=\theta)}$ model, uses the θ parameter to scale differences in travel cost within both the probability relation and the path size contribution factor, and thus for low θ Route 3 has the lowest choice probability due to being the least distinct, and for larger θ Route 3's choice probability converges to Route 2 as Route 4's path size contribution diminishes. Fig. 9B shows the example network 4 APSL choice probabilities for increasing θ, $\beta = 1.4$. An increase in β further decreases Route 3's choice probability for low θ, and as a consequence Route 3's path size contribution to Route 4 diminishes and Route 4 converges to the choice probability of Route 1.

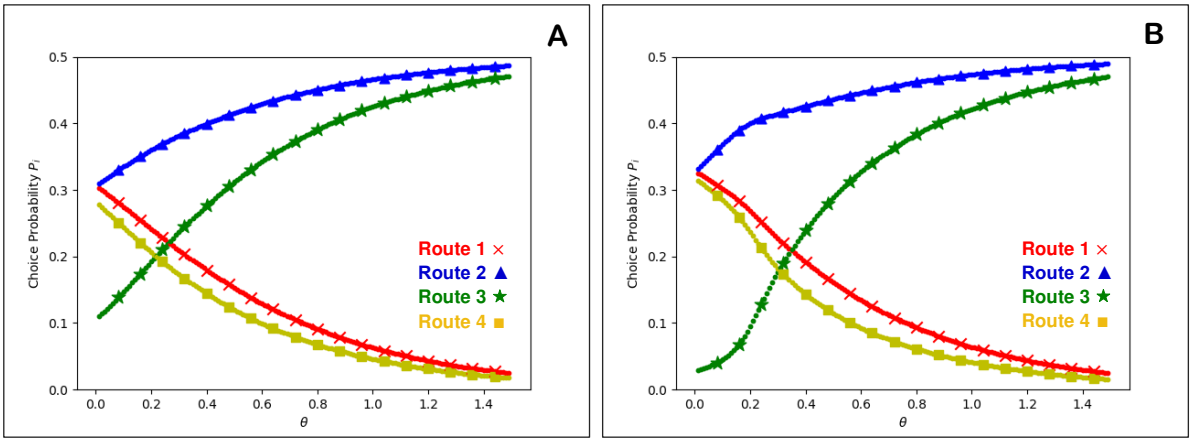


Fig. 9. Example network 4: APSL route choice probabilities for increasing θ – A: $\beta = 1$. B: $\beta = 1.4$.

3.4 Solution Method

There are many fixed-point algorithms available for solving the APSL fixed-point system $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$. In the studies in this paper we utilise the simplest fixed-point algorithm available: the Fixed-Point Iteration Method (FPIM)

(Isaacson & Keller, 1966). The FPIM is the most basic fixed-point algorithm, and other algorithms aim to accelerate the convergence of the FPIM, though require more complicated computations at each iteration. The FPIM for solving the APSL fixed-point system $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ is formulated as follows:

$$P_i^{(s+1)} = G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}\left(\boldsymbol{P}^{(s)}\right)\right)\right), \qquad s = 0,1,2,\dots$$

such that

$$\lim_{s\to\infty} P_i^{(s+1)} = \lim_{k\to\infty} G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}\left(\boldsymbol{P}^{(s)}\right)\right)\right) = P_i^*, \qquad \boldsymbol{P}^{(0)} \in D^{(\tau)}, \qquad \forall i \in R.$$

A standard convergence statistic we chose to observe in this study is $\ln\left(\sum_{i\in R}\left|P_i^{(s+1)} - P_i^{(s)}\right|\right)$, and the FPIM is said to have converged sufficiently to an APSL choice probability solution if:

$$\ln\left(\sum_{i\in R}\left|P_i^{(s+1)} - P_i^{(s)}\right|\right) < \ln\left(10^{-\xi}\right),$$

where $\xi > 0$ is a predetermined convergence parameter. In Sections 5.3.2.2 & 5.4.1.2 we assess the computational performance of the APSL model in calculating choice probabilities and parameter estimation.

4 Existence and Uniqueness of APSL Solutions

In this section we establish a series of theoretical results concerning the APSL model as defined in (11), (12), and (13), where the guaranteed existence of solutions is proven, and sufficient conditions for the uniqueness of solutions are detailed.

4.1 Properties

First, we note the relationship between the APSL$_0$ and APSL models.

Property 2. *A solution to the APSL model as defined in* (11), (12), *and* (13) *approaches the APSL$_0$ model as defined in* (7) *and* (8) *in the limit as* $\tau \to 0$.

Proof. This follows by inspection from the definition of G_i in (11) noting that $G_i \to g_i$, as $\tau \to 0$.

■

From Property 2, the APSL model will thus satisfy Property 1 in the limit as $\tau \to 0$.

We next provide two important properties of the fixed-point function $\boldsymbol{G}$. In Lemma 1 we establish the continuity property of $\boldsymbol{G}$.

Lemma 1. $G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ *is a continuous function for* $\boldsymbol{P} \in D^{(\tau)}$, $\forall i \in R$.

Proof. From the definition (13) above it follows that $\boldsymbol{\gamma}^{APS}$ is continuous in $\boldsymbol{P}$ for all $\boldsymbol{P} \in D^{(\tau)}$:

$$\lim_{\boldsymbol{P}\to\boldsymbol{q}} \boldsymbol{\gamma}^{APS}(\boldsymbol{P}) = \boldsymbol{\gamma}^{APS}(\boldsymbol{q}), \qquad \forall \boldsymbol{q} \in D^{(\tau)}. \tag{14}$$

If we let Γ be the set of possible path size terms:

$$\Gamma = \{\boldsymbol{\gamma}^{APS} \in \mathbb{R}^N_{>0} : 0 < \gamma_i^{APS} \leq 1, \forall i \in R\},$$

then from definition (12) above it follows that g_i is continuous in $\boldsymbol{\gamma}^{APS}$ for all $\boldsymbol{\gamma}^{APS} \in \Gamma$:

$$\lim_{\boldsymbol{\gamma}^{APS}\to\boldsymbol{\gamma}_0} g_i(\boldsymbol{\gamma}^{APS}) = g_i(\boldsymbol{\gamma}_0), \qquad \forall \boldsymbol{\gamma}_0 \in \Gamma, \qquad \forall i \in R. \tag{15}$$

And, from definition (11) above it follows that G_i is continuous in x for all $x \in (0,1)$:

$$\lim_{x\to x_0} G_i(x) = G_i(x_0), \qquad \forall x_0 \in (0,1). \tag{16}$$

It then follows from (14), (15) and (16) that $G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$, as a composition of continuous functions, is itself continuous in $\boldsymbol{P}$ for all $\boldsymbol{P} \in D^{(\tau)}$:

$$\lim_{\boldsymbol{P}\to\boldsymbol{q}} G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right) = G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{q})\right)\right), \qquad \forall \boldsymbol{q} \in D^{(\tau)}, \qquad \forall i \in R.$$

■

We now in Lemma 2 show that the domain of $\boldsymbol{G}$ maps to itself.

Lemma 2. $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ *maps* $D^{(\tau)}$ *into* $D^{(\tau)}$.

Proof. From definition (13) above it follows that the function $\boldsymbol{\gamma}^{APS}$ maps $D^{(\tau)} \to \Gamma$, from definition (12) it follows that the function $\boldsymbol{g}$ maps $\Gamma \to D^{>0}$, and, from definition (11) it follows that the function $\boldsymbol{G}$ maps $D^{>0} \to D^{(\tau)}$. It thus follows that the composition of the functions $\boldsymbol{\gamma}^{APS}$, $\boldsymbol{g}$, and $\boldsymbol{G}$, $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$, maps $D^{(\tau)} \to D^{(\tau)}$.

■

4.2 Existence of Solutions

Having established some properties regarding the APSL fixed-point function $\boldsymbol{G}$, we consider the existence of APSL solutions.

Proposition 1. *At least one APSL fixed-point solution, $\boldsymbol{P}^*$, to the system $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ is guaranteed to exist in $D^{(\tau)}$.*

Proof. $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ is a continuous function by Lemma 1, which maps $D^{(\tau)}$ into $D^{(\tau)}$ by Lemma 2, and thus since $D^{(\tau)}$ is a compact convex set, and by Brouwer's Fixed-Point Theorem at least one fixed-point solution, $\boldsymbol{P}^*$, is guaranteed to exist for the system $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)$ in $D^{(\tau)}$.

■

4.3 Uniqueness of Solutions

Having proven that APSL solutions are guaranteed to exist, the next question is whether sufficient conditions exist which ensure APSL solutions are unique. In order to do this, we must first establish two key properties of $J_{\boldsymbol{G}}(\boldsymbol{P};\beta)$ which is the Jacobian matrix of first partial derivatives of $\boldsymbol{G}$ evaluated at $\boldsymbol{P}$ and β.

Lemma 3. *The maximum Jacobian matrix norm of $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta\right)\right)$ for all $\boldsymbol{P} \in D^{(\tau)}$ at $\beta = 0$ is equal to zero:* $\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};0)\|: \forall \boldsymbol{P} \in D^{(\tau)}\right) = 0$.

Proof. From definitions (11) and (12) above it follows that:

$$G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});0\right)\right) = \tau + (1 - N\tau) \cdot \left(\frac{e^{-\theta c_i}}{\sum_{j \in R} e^{-\theta c_j}}\right), \qquad \forall i \in R. \tag{17}$$

It then follows from (17) that:

$$\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});0\right)\right)}{\partial P_l} = 0, \qquad \forall \boldsymbol{P} \in D^{(\tau)}, \qquad \forall i, l \in R. \tag{18}$$

It thus follows from (18) that $\|J_{\boldsymbol{G}}(\boldsymbol{P};0)\| = 0, \forall \boldsymbol{P} \in D^{(\tau)}$, and hence $\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};0)\|: \forall \boldsymbol{P} \in D^{(\tau)}\right) = 0$.

■

Lemma 4. *The maximum Jacobian matrix norm of $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta\right)\right)$, $\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\|: \forall \boldsymbol{P} \in D^{(\tau)}\right)$, is a continuous function for $\beta \in [0, \infty)$.*

Proof. It follows from the definitions (11), (12), and (13) above that:

$$\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)}{\partial P_l}$$

$$= (1 - N\tau) \cdot \left(\frac{\left(\gamma_i^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_i}}{\sum_{j \in R}\left(\gamma_j^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_j}}\right) \cdot \left(\frac{\beta \frac{\partial \gamma_i^{APS}(\boldsymbol{P})}{\partial P_l}}{\left(\gamma_i^{APS}(\boldsymbol{P})\right)} - \frac{\left(\sum_{j \in R} \beta \frac{\partial \gamma_j^{APS}(\boldsymbol{P})}{\partial P_l}\left(\gamma_j^{APS}(\boldsymbol{P})\right)^{\beta-1} e^{-\theta c_j}\right)}{\left(\sum_{j \in R}\left(\gamma_j^{APS}(\boldsymbol{P})\right)^{\beta} e^{-\theta c_j}\right)}\right), \qquad \forall i, l \in R, \tag{19}$$

$$\frac{\partial \gamma_i^{APS}(\boldsymbol{P})}{\partial P_i} = \sum\nolimits_{a \in A_i} \frac{t_a}{c_i}\left(\frac{\sum_{k \in R; k \neq i} P_k \delta_{a,k}}{\left(\sum_{k \in R} P_k \delta_{a,k}\right)^2}\right), \qquad \forall i \in R, \tag{20}$$

and,

$$\frac{\partial \gamma_i^{APS}(\boldsymbol{P})}{\partial P_l} = -\sum\nolimits_{a \in A_i} \frac{t_a}{c_i} \frac{P_i \delta_{a,l}}{\left(\sum_{k \in R} P_k \delta_{a,k}\right)^2}, \qquad \forall i, l \in R, l \neq i. \tag{21}$$

From the definitions (20) and (21) above it follows that $\frac{\partial \boldsymbol{\gamma}^{APS}(\boldsymbol{P})}{\partial \boldsymbol{P}}$ is continuous in $\boldsymbol{P}$ for all $\boldsymbol{P} \in D^{(\tau)}$:

$$\lim_{\boldsymbol{P}\to\boldsymbol{q}} \frac{\partial \boldsymbol{\gamma}^{APS}(\boldsymbol{P})}{\partial \boldsymbol{P}} = \frac{\partial \boldsymbol{\gamma}^{APS}(\boldsymbol{q})}{\partial \boldsymbol{P}}, \qquad \forall \boldsymbol{q} \in D^{(\tau)}. \tag{22}$$

It then follows from (14) and (22) that $\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)}{\partial P_l}$ as defined in (19), being a composition of continuous functions, is itself continuous in $\boldsymbol{P}$ for all $\boldsymbol{P} \in D^{(\tau)}$:

$$\lim_{\boldsymbol{P}\to\boldsymbol{q}} \frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)}{\partial P_l} = \frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{q})\right)\right)}{\partial P_l}, \qquad \forall \boldsymbol{q} \in D^{(\tau)}, \qquad \forall i, l \in R.$$

Since $\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P})\right)\right)}{\partial P_l}$ is a continuous function for $\boldsymbol{P} \in D^{(\tau)}$, $\forall i, l \in R$, $\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta\right)\right)}{\partial P_l}$ is also a continuous function for $\beta \in [0, \infty)$, $\forall i, l \in R$:

$$\lim_{\beta\to\beta_0} \left(\frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta\right)\right)}{\partial P_l} \right) = \frac{\partial G_i\left(g_i\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta_0\right)\right)}{\partial P_l}, \qquad \forall \beta_0 \in [0, \infty), \qquad \forall i, l \in R.$$

Hence, since $\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| : \forall \boldsymbol{P} \in D^{(\tau)}\right)$ is a composition of continuous functions then it itself is a continuous function for $\beta \in [0, \infty)$:

$$\lim_{\beta\to\beta_0} \left(\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| : \forall \boldsymbol{P} \in D^{(\tau)}\right)\right) = \max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta_0)\| : \forall \boldsymbol{P} \in D^{(\tau)}\right), \qquad \forall \beta_0 \in [0, \infty).$$

■

These two key properties of $J_{\boldsymbol{G}}(\boldsymbol{P};\beta)$ allow us to establish conditions for the uniqueness of solutions.

Proposition 2. *There always exist values of $b > 0$ such that when the β parameter is within the range $0 \le \beta \le b$, there are unique APSL fixed-point solutions, $\boldsymbol{P}^*$, to the system $\boldsymbol{P} = \boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{P});\beta\right)\right)$ in $D^{(\tau)}$.*

Proof. $\boldsymbol{G}$ is a *contraction mapping* on the domain $D^{(\tau)}$ if:

a) $\boldsymbol{G}$ maps $D^{(\tau)}$ into itself, so $\boldsymbol{G}\left(\boldsymbol{g}\left(\boldsymbol{\gamma}^{APS}(\boldsymbol{q});\beta\right)\right) \in D^{(\tau)}, \forall \boldsymbol{q} \in D^{(\tau)}$, and

b) There exists a constant $0 \le \sigma < 1$ such that:

$$\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| \le \sigma, \qquad \forall \boldsymbol{P} \in D^{(\tau)},$$

where $J_{\boldsymbol{G}}(\boldsymbol{P};\beta)$ is the Jacobian matrix of first partial derivatives of $\boldsymbol{G}$ evaluated at $\boldsymbol{P}$.

If the link cost vector $\boldsymbol{t}$ is fixed (and thus the route cost vector $\boldsymbol{c}$ is fixed), and θ is fixed, then for any given β, if $\boldsymbol{G}$ is a contraction mapping, then since $D^{(\tau)}$ is a compact convex set, and by Lemma 1, Lemma 2, and the Contraction Mapping Theorem, $\boldsymbol{G}$ emits a unique fixed-point solution $\boldsymbol{P}^* \in D^{(\tau)}$.

It remains to establish the conditions under which $\boldsymbol{G}$ is a contraction mapping. Since by Lemma 3 the maximum Jacobian matrix norm of $\boldsymbol{G}$ for all $\boldsymbol{P} \in D^{(\tau)}$ at $\beta = 0$ is equal to zero $\left(\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};0)\| : \forall \boldsymbol{P} \in D^{(\tau)}\right) = 0\right)$, and by Lemma 4 $\max\left(\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| : \forall \boldsymbol{P} \in D^{(\tau)}\right)$ is a continuous function for $\beta \in [0, \infty)$, then there must always exist values $b > 0$ such that when β is within the range $0 \le \beta \le b$ $\boldsymbol{G}$ is a contraction mapping and the sufficient conditions for unique APSL solutions are always met.

■

There are cases where the APSL model has unique solutions for all $\beta > 0$ (i.e. for all values of b), for example where all routes are non-overlapping and the path size terms are consequently all 1 so that G_i collapses to (17), and hence in these cases a maximum value for b does not exist. However, in most cases APSL solutions are not unique for all values of β and in these cases a maximum value for b exists (denoted b_{max}) such that Proposition 2 holds. However, Proposition 2 is only a sufficient condition for unique APSL solutions and solutions are not necessarily non-unique for $\beta > b_{max}$. In Section 4.4 below we explore how $0 \le \beta \le b_{max}$ relates to the true maximum range $0 \le \beta \le \beta_{max}$ in which APSL solutions are unique.

4.4 Investigating the Conditions for Uniqueness

b_{max} is the maximum value such that the sufficient conditions for unique APSL solutions in Proposition 2 are satisfied for all β in the range $0 \le \beta \le b_{max}$. β_{max} is the true maximum value such that APSL solutions are unique for all β in the range $0 \le \beta \le \beta_{max}$, where $\beta_{max} \ge b_{max}$. The purpose of this section is to explore how close b_{max} is to β_{max}, and demonstrate that multiple solutions can exist when β is greater than β_{max}. We specify and demonstrate how to calculate b_{max}, and suggest and demonstrate a method for estimating β_{max}.

Calculating b_{max} can be formulated as either of the following optimisation problems:

b_{max} Optimisation Problem 1

$$b_{max} = \max\{\beta\}$$

subject to

$$\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| < 1, \qquad \forall \boldsymbol{P} \in D^{(\tau)}.$$

b_{max} Optimisation Problem 2

$$b_{max} = \max\{\beta\}$$

subject to

$$\|J_{\boldsymbol{G}}(\bar{\boldsymbol{P}};\beta)\| < 1$$

where

$$\bar{\boldsymbol{P}} = \arg\max_{\boldsymbol{P}}\{\|J_{\boldsymbol{G}}(\boldsymbol{P};\beta)\| : \forall \boldsymbol{P} \in D^{(\tau)}\}.$$

Example

Consider example network 5 in Fig. 10 where there are 2 routes: Route 1 has travel cost $u + w$ and Route 2 has travel cost $v + w$. Fig. 11A-D display the maximum Jacobian matrix norm of $\boldsymbol{G}$ for increasing β for four different network parameter settings.

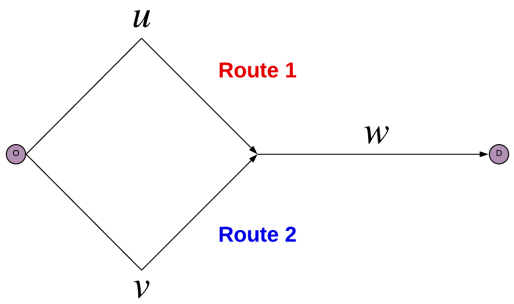
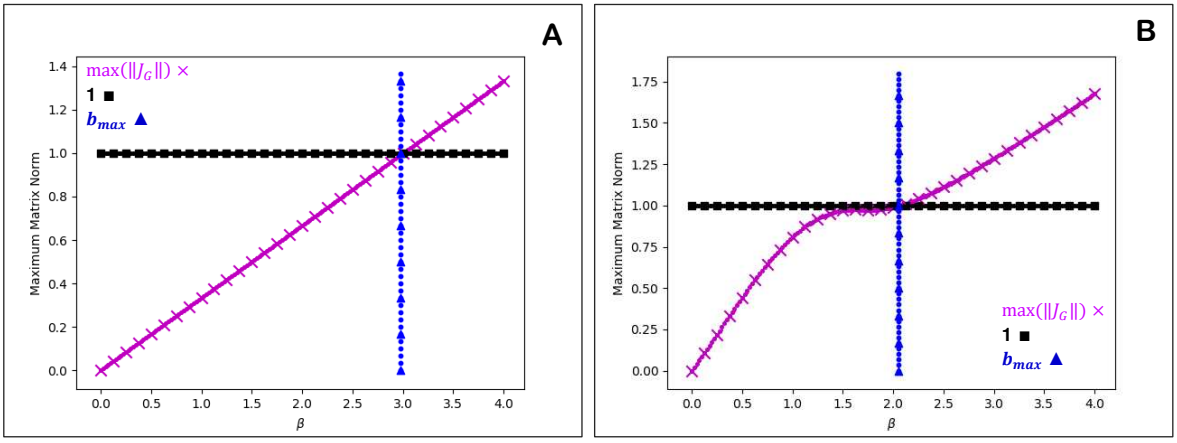


Fig. 10. Example network 5.



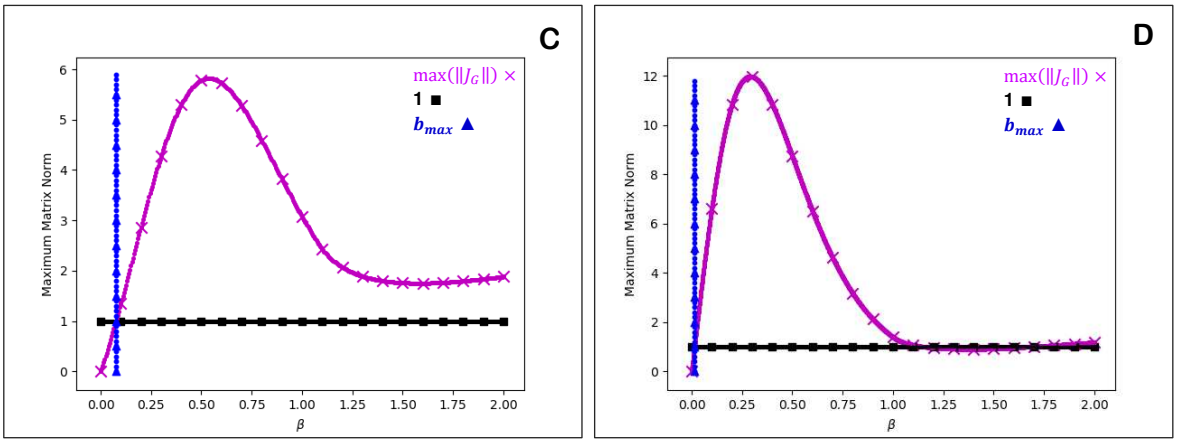


Fig. 11. Example network 5: Maximum Jacobian matrix norm of APSL fixed-point function $\boldsymbol{G}$ for increasing β – A: $u = v = w = \theta = 1$, ($b_{max} \cong 2.95$). B: $u = 0.5$, $v = w = \theta = 1$, ($b_{max} \cong 2.05$). C: $u = 0.1$, $v = 1$, $w = 4$, $\theta = 1$, ($b_{max} \cong 0.075$). D: $u = 0.1$, $v = w = 20$, $\theta = 10^{-5}$, ($b_{max} \cong 0.0125$).

β_{max} can be estimated by plotting trajectories of APSL solutions for varying β, and identifying where a unique trajectory of solutions ends and multiple trajectories begin. We briefly detail here a simple method for obtaining trajectories of APSL solutions:

Step 1. Identify a suitably large value for β.

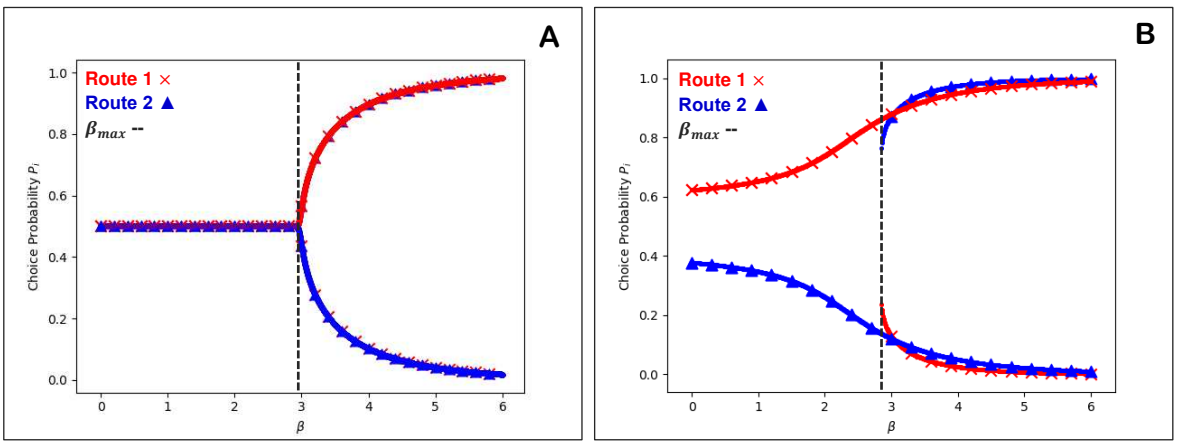
Step 2. Solve the APSL fixed-point system for this large β with a randomly generated initial condition (see Section 3.4).

Step 3. Decrement β and obtain the next APSL solution with initial condition set as the APSL solution for the previous β.

Step 4. Continue until $\beta = 0$ or $\beta = b_{max}$ if known.

By plotting the choice probabilities at each decremented β, and repeating this method several times, one can determine where non-unique solution trajectories end and hence estimate β_{max}. If after several repetitions (with different randomly generated initial conditions) only a single trajectory of solutions is shown, then the initial large β value is increased. We illustrate the approach graphically, but there is no need to draw graphs for general networks. One can instead observe the choice probability values, where a finer grained decrement of β will provide a more accurate estimation of β_{max}.

To demonstrate, consider again example network 5 in Fig. 10; Fig. 12A-D display trajectories of APSL solutions as the β parameter is varied for the same network parameter settings as Fig. 11A-D, respectively. β was decremented by 0.01, and the initial large β value was $\beta = 10$, (though trajectories are only plotted for part of this range for illustrative purposes). The solution trajectory plotting was repeated until multiple clear trajectories were shown. As each of Fig. 12A-D show, there is a unique trajectory of choice probability solutions up until $\beta = \beta_{max}$ where there then becomes multiple trajectories. As Fig. 12A shows, the estimated β_{max} value in this case is equal to b_{max} in Fig. 11A, however in the cases of Fig. 12B-D, the estimated β_{max} values are all greater than the b_{max} values in Fig. 11B-D.



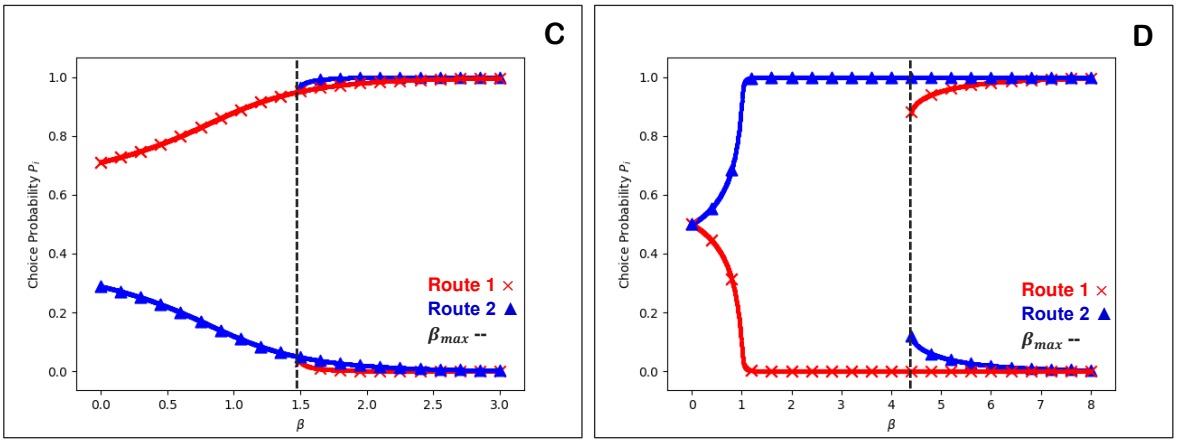


Fig. 12. Example network 5: Trajectories of APSL choice probability solutions as β is varied – A: $u = v = w = \theta = 1$, ($\beta_{max} \cong 2.95$). B: $u = 0.5$, $v = w = \theta = 1$, ($\beta_{max} \cong 2.8$). C: $u = 0.1$, $v = 1$, $w = 4$, $\theta = 1$, ($\beta_{max} \cong 1.5$). D: $u = 0.1$, $v = w = 20$, $\theta = 10^{-5}$, ($\beta_{max} \cong 4.3$).

As we have shown, b_{max} can be a conservative estimate of β_{max}. On the other hand, we have shown that for large enough values of β multiple solutions can exist, and therefore is an issue that should be considered in practice in real-life applications. The experiments above provide a computationally feasible method for revealing multiple solutions, and thus β_{max}, that can be applied in realistic sized networks (see Sections 5.3.2.3 / 5.4.1.3).

5 Estimating the APSL Model

In this section, we provide a Maximum Likelihood Estimation (MLE) procedure for estimating the APSL model with tracked route observations. This procedure for estimating the APSL model is then investigated in a simulation study, and the possibility of retrieving APSL parameter estimates is assessed. The APSL model is then estimated on a large-scale network using real route choice observation data tracked with GPS units, and results are compared with other Path Size Logit models.

5.1 Notation and Definitions for Estimation with Multiple OD Movements

5.1.1 Notation

To consider the estimation of the APSL model and other Path Size Logit models, we extend definitions here for estimation on a network with multiple OD movements, but where the travel costs remain fixed. The road network consists of link set A and $m = 1, \dots, M$ OD movements. R_m is the choice set of all simple routes (no cycles) for OD movement m of size $N_m = |R_m|$, and $A_{m,i} \subseteq A$ is the set of links belonging to route $i \in R_m$, and $\delta_{a,m,i} = \begin{cases} 1 & if\ a \in A_{m,i} \\ 0 & otherwise \end{cases}$. Suppose that the generalised travel cost t_a of each link $a \in A$ is a weighted sum (by parameter vector $\boldsymbol{\alpha}$) of variables $\boldsymbol{w}_a$, i.e. $t_a(\boldsymbol{w}_a; \boldsymbol{\alpha})$, and that the generalised travel cost for route $i \in R_m$, $c_{m,i}$, can be attained through summing up the total cost of its links so that $c_{m,i}\big(\boldsymbol{t}(\boldsymbol{w}; \boldsymbol{\alpha})\big) = \sum_{a \in A_{m,i}} t_a(\boldsymbol{w}_a; \boldsymbol{\alpha})$, where $\boldsymbol{t}$ is the vector of all link travel costs and $\boldsymbol{w}$ is the vector of all link variables. Let the route choice probability for route $i \in R_m$ be $P_{m,i}$, where $\boldsymbol{P}_m = \big(P_{m,1}, P_{m,2}, \dots, P_{m,N_m}\big)$ is the vector of route choice probabilities for OD movement m, and D_m is the domain of possible route choice probability vectors for OD movement m, $m = 1, \dots, M$.

5.1.2 Model Definitions

5.1.2.1 Multinomial Logit

MNL choice probability relation for route $i \in R_m$:

$$P_{m,i}(\boldsymbol{t}) = \frac{1}{\sum_{j \in R_m} e^{-\theta\left(c_{m,j}(\boldsymbol{t}) - c_{m,i}(\boldsymbol{t})\right)}}. \tag{23}$$

5.1.2.2 Regular Path Size Logit Models

Regular Path Size Logit model choice probability relation for route $i \in R_m$:

$$P_{m,i}(\boldsymbol{t}) = \frac{1}{\sum_{j\in R_m}\left(\frac{\gamma_{m,j}(\boldsymbol{t})}{\gamma_{m,i}(\boldsymbol{t})}\right)^{\beta} e^{-\theta\left(c_{m,j}(\boldsymbol{t})-c_{m,i}(\boldsymbol{t})\right)}}. \tag{24}$$

PSL path size term for route $i \in R_m$:

$$\gamma_{m,i}^{PS}(\boldsymbol{t}) = \sum_{a\in A_{m,i}} \frac{t_a}{c_{m,i}(\boldsymbol{t})} \frac{1}{\sum_{k\in R_m}\delta_{a,m,k}}. \tag{25}$$

GPSL path size term for route $i \in R_m$:

$$\gamma_{m,i}^{GPS}(\boldsymbol{t}) = \sum_{a\in A_{m,i}} \frac{t_a}{c_{m,i}(\boldsymbol{t})} \frac{1}{\sum_{k\in R_m}\left(\frac{c_{m,i}(\boldsymbol{t})}{c_{m,k}(\boldsymbol{t})}\right)^{\lambda}\delta_{a,m,k}}. \tag{26}$$

GPSL$'_{(\lambda=\theta)}$ path size term for route $i \in R_m$:

$$\gamma_{m,i}^{GPS'}(\boldsymbol{t}) = \sum_{a\in A_{m,i}} \frac{t_a}{c_{m,i}(\boldsymbol{t})} \frac{1}{\sum_{k\in R_m} e^{-\theta\left(c_{m,k}(\boldsymbol{t})-c_{m,i}(\boldsymbol{t})\right)}\delta_{a,m,k}}. \tag{27}$$

5.1.2.3 Adaptive Path Size Logit

The APSL choice probabilities for OD movement m, $\boldsymbol{P}_m^*$, are a solution to the fixed-point problem $\boldsymbol{P}_m = \boldsymbol{G}_m\left(\boldsymbol{g}_m\left(\boldsymbol{c}_m(\boldsymbol{t}), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right)\right)$, where:

$$G_{m,i}\left(g_{m,i}\left(\boldsymbol{c}_m(\boldsymbol{t}), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right)\right) = \tau_m + (1 - N_m\tau_m)\cdot g_{m,i}\left(\boldsymbol{c}_m(\boldsymbol{t}), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right), \tag{28}$$

$$g_{m,i}\left(\boldsymbol{c}_m(\boldsymbol{t}), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right) = \frac{\left(\gamma_{m,i}^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right)^{\beta} e^{-\theta c_{m,i}(\boldsymbol{t})}}{\sum_{j\in R_m}\left(\gamma_{m,j}^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right)^{\beta} e^{-\theta c_{m,j}(\boldsymbol{t})}}, \tag{29}$$

$$\gamma_{m,i}^{APS}(\boldsymbol{t}, \boldsymbol{P}_m) = \sum_{a\in A_{m,i}} \frac{t_a}{c_{m,i}(\boldsymbol{t})} \frac{P_{m,i}}{\sum_{k\in R_m} P_{m,k}\delta_{a,m,k}}, \tag{30}$$

$$\forall i \in R_m, \qquad \forall \boldsymbol{P}_m \in D_m^{(\tau_m)}, \qquad 0 < \tau_m \leq \frac{1}{N_m},$$

$$D_m^{(\tau_m)} = \left\{\boldsymbol{P}_m \in \mathbb{R}_{>0}^{N_m} : \tau_m \leq P_{m,i} \leq (1 - (N_m - 1)\tau_m), \forall i \in R_m, \sum_{j=1}^{N_m} P_{m,j} = 1\right\},$$

$$m = 1, \dots, M, \qquad \theta > 0, \qquad \beta \geq 0.$$

Each OD movement has its own range restrictions for τ_m based on the number of routes in the choice set, but the τ_m parameters are not model parameters that require estimating, they are simply a mathematical construct that ensure solutions to the APSL model exist and can be unique. As discussed in Section 3.2, only small values of τ_m should be used, and we set $\tau_m = 10^{-16}$, $m = 1, \dots, M$. APSL choice probability solutions are computed using the FPIM with initial conditions set as the MNL route choice probabilities, and convergence statistic set at $\xi = 10$ (see Section 3.4), unless specified otherwise.

5.2 Adaptive Path Size Logit Likelihood Formulation & Estimation Procedure

5.2.1 Likelihood Formulation

Suppose that we have available a set of Z observed routes, e.g. collected through GPS units or smart phones, and consider a situation where it is not needed to distinguish individuals in their preferences (the approach is, of course, readily generalised to permit multiple user classes differing in their parameters). Let m_z denote the OD movement of route observation z, and for each trip observation $z = 1,2, \dots, Z$, let R_{m_z} be the choice set of all simple routes between the origin and destination of the trip. Suppose that the observation data is contained in a vector $\boldsymbol{x}$ of size Z where:

$$x_z = i \quad \text{if alternative } i \in R_{m_z} \text{ is chosen}, \qquad z = 1, \dots, Z.$$

The Likelihood for a sample of size Z, can be formulated as:

$$L(\boldsymbol{\alpha}, \theta, \beta | \boldsymbol{x}) = \prod_{z=1}^{Z} P^*_{m_z, x_z}(\boldsymbol{t}(\boldsymbol{w}; \boldsymbol{\alpha}), \theta, \beta)\,, \tag{31}$$

where $P^*_{m_z,x_z}(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha}),\theta,\beta)$ is the APSL choice probability solution for route $x_z \in R_{m_z}$ to the fixed-point problem $\boldsymbol{P}_{m_z} = \boldsymbol{G}_{m_z}\left(\boldsymbol{g}_{m_z}\left(\boldsymbol{c}_{m_z}(\boldsymbol{t}), \boldsymbol{\gamma}^{APS}_{m_z}(\boldsymbol{t}, \boldsymbol{P}_{m_z})\right)\right)$ for OD movement m_z, where $G_{m,i}$ and $g_{m,i}$ are as in (28) and (29) for route $i \in R_m$, respectively. The Log-Likelihood function is thus:

$$LL(\boldsymbol{\alpha}, \theta, \beta | \boldsymbol{x}) = \ln\left(\prod_{z=1}^{Z} P^*_{m_z, x_z}(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha}), \theta, \beta)\right) = \sum_{z=1}^{Z} \ln\left(P^*_{m_z, x_z}(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha}), \theta, \beta)\right) \tag{32}$$

5.2.2 Estimation Procedure

Standard Maximum Likelihood Estimation (MLE) procedures can be used to estimate the parameters of the APSL model for a given network. Using a standard iterative estimation procedure, APSL model parameters can be found that maximise the Log-Likelihood function as formulated in (32) above for a given set of data. Algorithm 1 outlines pseudo-code for the estimation procedure.

Step 1: Initialisation. For each route observation $z = 1, \dots, Z$, generate the corresponding universal choice set and store the link attributes and link-route information. Define an initial set of parameter values $\left(\tilde{\boldsymbol{\alpha}}^{(1)}, \tilde{\theta}^{(1)}, \tilde{\beta}^{(1)}\right)$ for MLE, and set $n = 1$.

Step 2: Recalculate choice probabilities and LL. Given the set of parameter values $\left(\tilde{\boldsymbol{\alpha}}^{(n)}, \tilde{\theta}^{(n)}, \tilde{\beta}^{(n)}\right)$ for iteration n, calculate the link costs $\boldsymbol{t}\left(\boldsymbol{w}; \tilde{\boldsymbol{\alpha}}^{(n)}\right)$ and solve each of the fixed-point problems

$$\boldsymbol{P}_{m_z} = \boldsymbol{G}_{m_z}\left(\boldsymbol{g}_{m_z}\left(\boldsymbol{c}_{m_z}\left(\boldsymbol{t}\left(\boldsymbol{w}; \tilde{\boldsymbol{\alpha}}^{(n)}\right)\right), \boldsymbol{\gamma}^{APS}_{m_z}\left(\boldsymbol{t}\left(\boldsymbol{w}; \tilde{\boldsymbol{\alpha}}^{(n)}\right), \boldsymbol{P}_{m_z}\right); \tilde{\theta}^{(n)}, \tilde{\beta}^{(n)}\right)\right)$$

for $z = 1, \dots, Z$. Given the fixed-point choice probability solutions $P^*_{m_z,x_z}$ for each of the route observations $z = 1, \dots, Z$, calculate the Log-Likelihood $LL^{(n)}\left(\tilde{\boldsymbol{\alpha}}^{(n)}, \tilde{\theta}^{(n)}, \tilde{\beta}^{(n)} | \boldsymbol{x}\right)$ for iteration n.

Step 3: Compute new set of parameters. Based on $LL^{(s)}$ and the associated parameters $\left(\tilde{\boldsymbol{\alpha}}^{(s)}, \tilde{\theta}^{(s)}, \tilde{\beta}^{(s)}\right)$ for all $s \leq n$, compute a new set of parameters $\left(\tilde{\boldsymbol{\alpha}}^{(n+1)}, \tilde{\theta}^{(n+1)}, \tilde{\beta}^{(n+1)}\right)$ to test in the following iteration.

Step 4: Stopping criteria. If $\left|LL^{(n)} - LL^{(n-1)}\right| < \zeta$, stop. Otherwise, set $n = n + 1$ and return to Step 2.

Algorithm 1: Pseudo-code for estimating the APSL model.

In general, Step 3 could apply procedures from standard numerical optimisation methods to identify the parameters to evaluate in the next iteration. Utilising gradient approaches such as Newton-Raphson or BHHH, however, is complicated by the difficulties in differentiating the APSL Log-Likelihood function, which involves differentiating the fixed-point choice probabilities with respect to the parameters, which is not straightforward. Other optimisation algorithms such as BFGS and alternative quasi-Newton algorithms use methods to approximate the differentials, and while are more computationally burdensome and typically less accurate, are readily useable. For the experiments in this paper, we utilise the L-BFGS-B bound-constraint, quasi-Newton minimisation algorithm (Byrd et al, 1995) for Steps 2-4 of Algorithm 1 (where we minimise $-LL$). The parameter bounds and initial conditions are given in each study.

Note that Algorithm 1 computes one set of parameter estimates, estimated from one set of observations. It is possible to calculate standard errors for the estimated APSL parameters analytically, but this is again complicated by the requirement to differentiate the APSL Log-Likelihood function with respect to the parameters. Instead, the robustness of the parameters estimated (variation of the estimates) can be investigated numerically by applying Algorithm 1 multiple times through resampling-approaches such as Bootstrap or Jackknife.

5.3 Simulation Study

In this section we investigate the formulated likelihood function for the APSL route choice model in a simulation study, evaluating the likelihood-surface and assessing the possibility of estimating reasonable parameters that reproduces observed behaviour.

5.3.1 Experiment Setup

In general, the approach is to sample observations according to an assumed 'true' model, and then use these in combination with the log-likelihood function to evaluate the ability to reproduce the assumed 'true' parameters. The simulation study consists of three steps:

(i) Postulate a true APSL choice model including specification and parameter values.

(ii) Sample a set of observed route choices according to the true model using the specified link travel costs.

(iii) Apply MLE approach to obtain parameter estimates based on the observed route choices.

The estimation procedure in Algorithm 1 is altered for simulation studies, by modifying **Step 1: Initialisation** as outlined below to reflect (i) and (ii) in the above:

Step 1: Initialisation.

1.1 For OD movements $m = 1, \dots, M$, generate the choice sets and store the link attributes and link-route information.

1.2. Postulate a true set of parameters $(\boldsymbol{\alpha}^{true}, \theta_{true}, \beta_{true})$ for the APSL model, and given these parameters and the generated choice sets, solve each of the fixed-point problems

$$\boldsymbol{P}_m = \boldsymbol{G}_m\left(\boldsymbol{g}_m\left(\boldsymbol{c}_m\left(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha}^{true})\right), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha}^{true}), \boldsymbol{P}_m); \theta_{true}, \beta_{true}\right)\right)$$

for $m = 1, \dots, M$.

1.3. Based on the fixed-point choice probability solutions $\boldsymbol{P}_m^*$ for $m = 1, \dots, M$ (obtained in 1.2), sample Z observed routes.

1.4. Define an initial set of parameter values $\left(\tilde{\boldsymbol{\alpha}}^{(1)}, \tilde{\theta}^{(1)}, \tilde{\beta}^{(1)}\right)$ for MLE, and set $n = 1$.

Algorithm 1 (Step 1): Pseudo-code, initialisation of simulation experiments.

The number of observed routes to sample, Z, is exogenously defined. The robustness of the estimated parameters estimated can be investigated numerically by applying Algorithm 1 multiple times and then analysing the variation of the estimated parameters.

5.3.2 Sioux Falls Application

The Sioux Falls network in Fig. 13 consists of 76 links, 528 OD movements with non-zero travel demands, and 1632820 total routes. The travel cost of link a is specified as the free-flow travel time $w_{a,1}$ only, such that:

$$t_a(\boldsymbol{w}_a;\boldsymbol{\alpha}) = w_{a,1} \cdot \alpha_1,$$

where $\alpha_1 > 0$ is the free-flow travel time parameter, and thus the travel cost for route $i \in R_m$ is:

$$c_{m,i}\left(\boldsymbol{t}(\boldsymbol{w};\boldsymbol{\alpha})\right) = \sum_{a \in A_{m,i}} t_a(\boldsymbol{w}_a;\boldsymbol{\alpha}) = \alpha_1 \sum_{a \in A_{m,i}} w_{a,1}.$$

The model requires the specification of three parameters: α_1, θ, and β, but to ensure identification θ is fixed at $\theta = 1$ throughout.

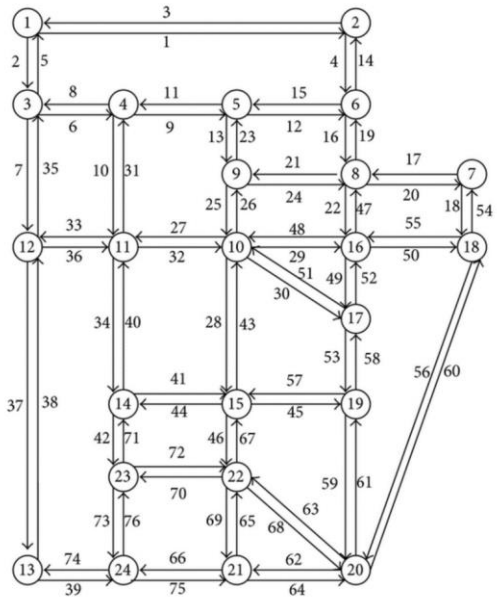


Fig. 13. Sioux Falls network.

Since the travel costs of the links (and thus routes) correspond to a single variable, to generate the utilised choice sets we employ k-shortest path to generate 150 of the lowest costing routes for each choice set. We also remove the short trip OD movements where the cheapest route has a free-flow travel time cost less than 10. The result is that there are 316 remaining OD movements, and a total of 47400 routes.

5.3.2.1 Experiment Results

We begin the simulation study by investigating the Log-Likelihood surface. By evaluating the Log-Likelihood function in (32) for various configurations of α_1 and β the Log-Likelihood surface can be visualised for a sample of observed routes. Fig. 14 displays the log-likelihood surface for a single estimation experiment, with $\alpha_1^{true} = 0.3$, $\beta_{true} = 0.6$, and $Z = 2000$. As Fig. 14 shows, the Log-Likelihood surface is smooth and approximately maximal around the true parameters, where the estimated parameters are $\hat{\alpha}_1 = 0.294 \pm 0.002$ and $\hat{\beta} = 0.57 \pm 0.01$.

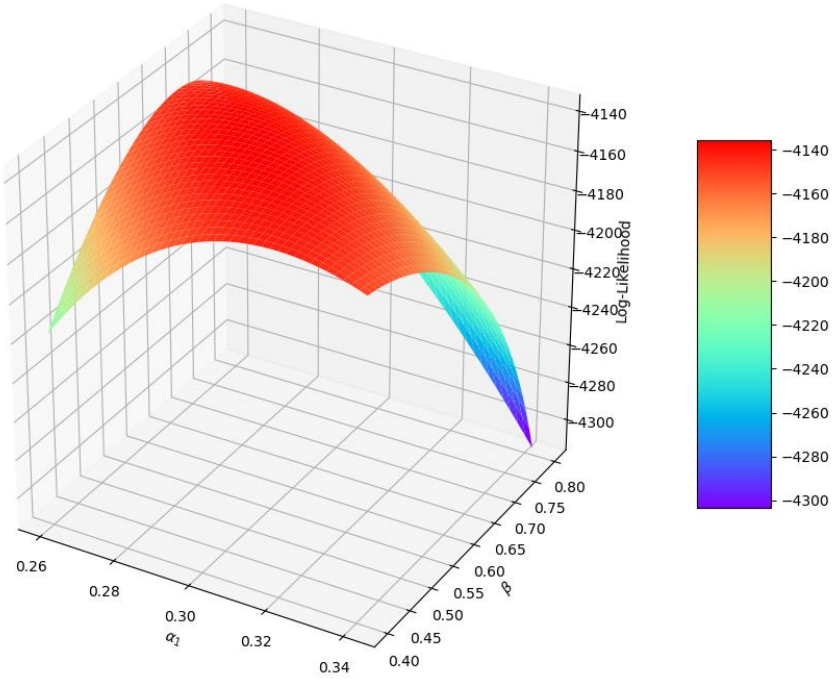


Fig. 14. Sioux Falls simulation study: Log-Likelihood surface; $\theta_{true} = 0.3$, $\beta_{true} = 0.6$, $Z = 2000$.

Next, we investigate the stability of the estimated parameters over multiple experiment replications. Each experiment utilises a Log-Likelihood maximisation algorithm (see Section 5.2.2) to obtain the parameter estimates with initial conditions $(\tilde{\alpha}_1^{(0)}, \tilde{\beta}^{(0)}) = (0.15,0)$, and bounds $\tilde{\alpha}_1, \tilde{\beta} \in [0,1]$.

Fig. 15A-D display for different settings of α_1^{true} and β_{true}, the distribution of the estimated parameters after $q = 100$ experiment replications of $Z = 1000$ simulated observations.

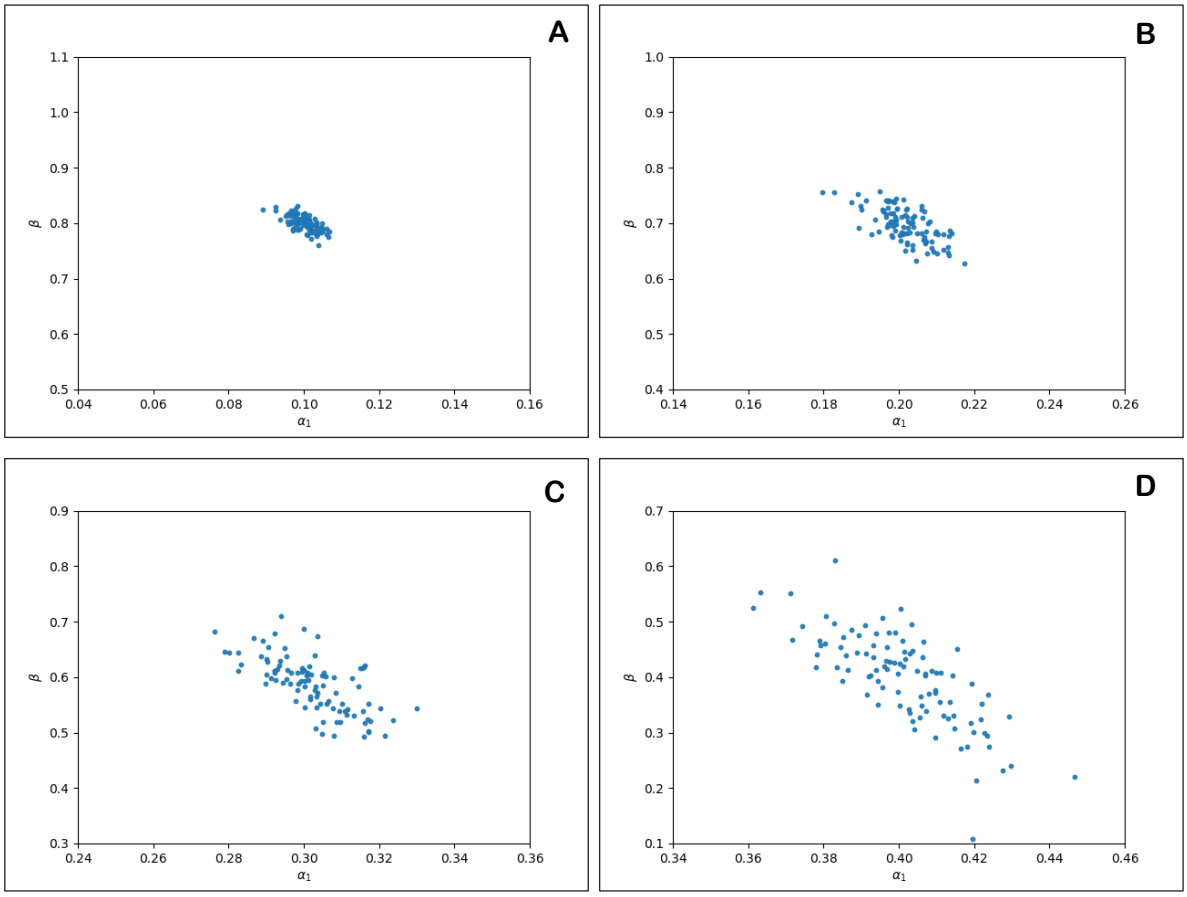


Fig. 15. Sioux Falls simulation study: Distribution of estimated parameters after multiple experiment replications; $Z = 1000$, $q = 100$ – A: $\alpha_1^{true} = 0.1$ $\beta_{true} = 0.8$. B: $\alpha_1^{true} = 0.2$ $\beta_{true} = 0.7$. C: $\alpha_1^{true} = 0.3$ $\beta_{true} = 0.6$. B: $\alpha_1^{true} = 0.4$ $\beta_{true} = 0.4$.

Table 2 reports, for various settings of the true parameters (same as Fig. 15A-D), the mean value (μ), standard deviation (σ), and Mean Squared Error (MSE) of the estimates across $q = 100$ experiment replications with $Z = 1000$ simulated observations. Table 2 also displays the estimated covariance between the α_1 and β parameters. As shown, the mean estimates of α_1 and β are close to the true values for all settings tested (i.e. there is no evidence of bias in the parameter estimates). However, as measured by the MSE, the precision of estimating α_1 decreases as α_1^{true} increases, and the precision of estimating β decreases as β_{true} decreases. This seems reasonable as increasing α_1 in this case corresponds to lower perception error of travel cost and decreasing β corresponds to lower perception of distinctiveness.

Table 2 and Fig. 17A-D both indicate that, with this network and the generated choice sets, there appears to be some negative correlation between the $\hat{\alpha}_1$ and $\hat{\beta}$ estimates. This is likely due to the large number of unrealistic routes present within the choice sets and the consequent small path size contribution factors for these routes from the true parameters; these factors can be reduced by increasing α_1 or β and hence negative correlation appears from balancing the parameters to obtain small contributions.

α_1^{true}	β_{true}	$\hat{\alpha}_1$			$\hat{\beta}$			$cov(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}})$
		μ	σ	MSE	μ	σ	MSE	
0.1	0.8	0.1000	0.0032	0.0000	0.8003	0.0136	0.0002	-0.00003
0.2	0.7	0.2016	0.0068	0.0001	0.6972	0.0305	0.0009	-0.00014
0.3	0.6	0.3021	0.0105	0.0001	0.5876	0.0485	0.0025	-0.00034
0.4	0.4	0.4012	0.0153	0.0002	0.3984	0.0817	0.0067	-0.00096

Table 2. Sioux Falls simulation study: Stability of estimated parameters across multiple experiment replications; $Z = 1000$, $q = 100$.

5.3.2.2 Computation Analysis

In this subsection we analyse the computational performance of the APSL model in the Sioux Falls MLE application. The computer used has a 2.10GHz Intel Xeon CPU, 512GB RAM, and 64 Logical Processors (of which 50 were utilised). The code was implemented in Python. Results are reported throughout this section for a single simulation experiment where $Z = 1000$ route choice observations were simulated from the true model $\alpha_1^{true} = 0.3$, $\beta_{true} = 0.6$. $\hat{\alpha}_1 = 0.2951$ and $\hat{\beta} = 0.6112$ are the consequent maximum likelihood estimates.

Fig. 16A shows for different values of the APSL choice probability convergence parameter ξ (and thus convergence statistic, See Section 3.4), the average number of fixed-point iterations per OD movement and computation time required to solve all of the 316 APSL fixed-point problems $\boldsymbol{P}_m = \boldsymbol{G}_m\left(\boldsymbol{g}_m\left(\boldsymbol{c}_m(\boldsymbol{t}), \boldsymbol{\gamma}_m^{APS}(\boldsymbol{t}, \boldsymbol{P}_m)\right)\right)$, and consequently compute the Log-Likelihood value of the maximum likelihood estimates. As shown, computation time and average number of fixed-point iterations per OD increase roughly linearly as the convergence parameter is increased. As expected, computation times relate to the number of iterations required for convergence. Fig. 16B shows the value of the Log-Likelihood obtained as ξ is increased. As shown, the Log-Likelihood increases in value and accuracy as the APSL choice probabilities become more accurate.

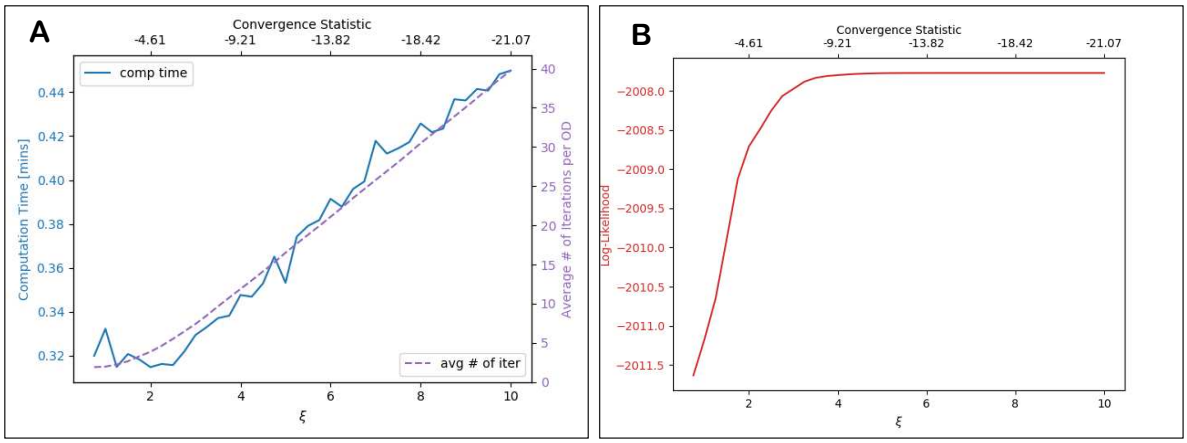


Fig. 16. Sioux Falls simulation study: Computational statistics for calculating APSL Log-Likelihoods as the APSL choice probability convergence parameter ξ is increased –
A: Average number of fixed-point iterations per OD / computation time [mins]. B: Log-Likelihood value.

Fig. 17 shows for different values of $\tilde{\beta}$ the average number of fixed-point iterations per OD movement and computation time required to calculate the Log-Likelihood, with $\xi = 7$ and $\tilde{\alpha}_1$ set as the maximum likelihood estimate $\hat{\alpha}_1 = 0.2965$. As shown, the average number of iterations per OD required for convergence increases as $\tilde{\beta}$ increases, and thus so do the required computation times.

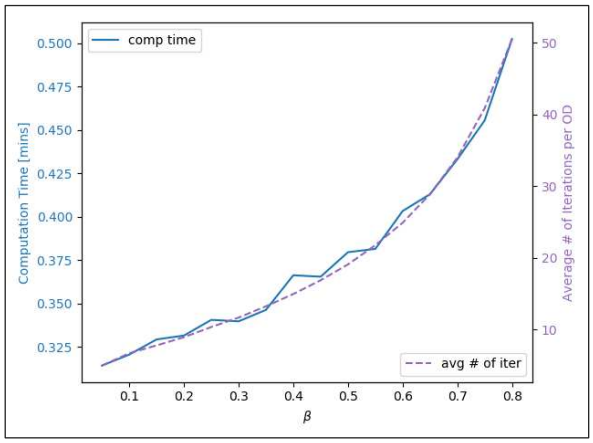


Fig. 17. Sioux Falls simulation study: Average number of fixed-point iterations per OD movement and computation time required to calculate the Log-Likelihood for different $\tilde{\beta}$ values; $\tilde{\alpha}_1 = \hat{\alpha}_1 = 0.2965$, $\xi = 7$.

Fig. 18A-B show for a single MLE (implementation of the L-BFGS-B algorithm), the cumulative computation times of the iterations and the Log-Likelihood values and parameter estimates at the end of each iteration, with $\xi = 7$.

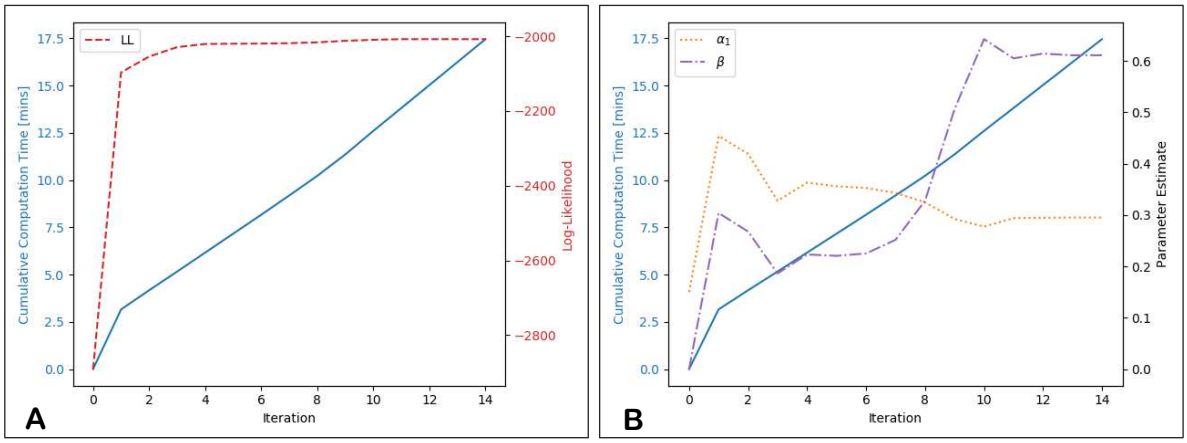


Fig. 18. Sioux Falls simulation study: Cumulative computation time at each iteration of a single MLE, and MLE statistics; $\xi = 7$ – A: Log-Likelihood. B: Parameter estimates.

Fig. 19A shows the total computation times and MLE final Log-Likelihood values of different MLE runs for different settings of ξ. Fig. 19B shows the parameter estimates. For $\xi = 5$, the inaccuracy of the APSL choice probabilities means that more MLE iterations are required to identify the estimates. For $\xi = 6$, however, the choice probabilities are sufficiently accurate to quickly estimate the parameters. Greater values of ξ increase the number of fixed-point iterations required for the fixed-point convergences and hence the computation times of the MLE.

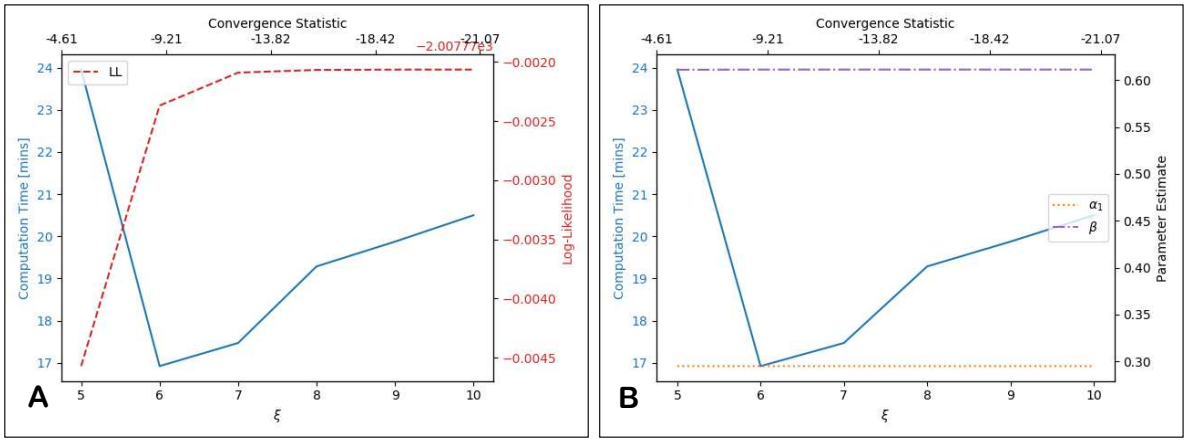


Fig. 19. Sioux Falls simulation study: Total computation time of MLE runs for different values of ξ, and MLE results – A: Final Log-Likelihood. B: Parameter estimates.

We also briefly investigate the impact of the τ parameter upon parameter estimation. Each OD movement m has a choice set size $N_m = 150$, and therefore supposing each OD movement has the same value for τ, the maximum value for τ is $\frac{1}{150} = 0.00\dot{6}$. Supposing τ assumes the form $\tau = 10^{-\varphi}$, Fig. 20 displays how the maximum likelihood parameter estimates vary as φ varies. As shown, the parameter estimates converge quickly to the limit case of $\tau \to 0$, demonstrating that we can recover the desired model APSL_0 (Section 3.1) to a high computational accuracy using the APSL model as defined in Section 3.2, with a sufficiently small value of τ.

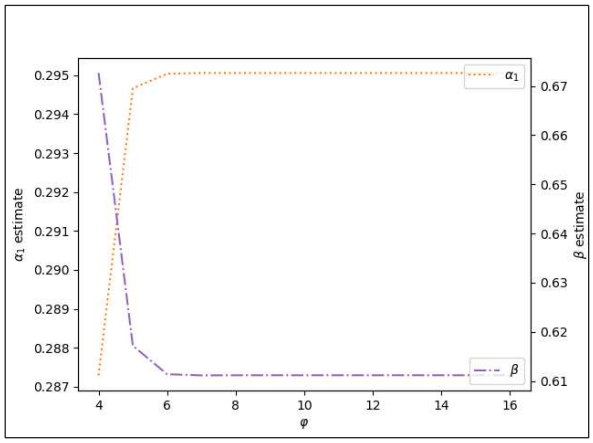


Fig. 20. Sioux Falls simulation study: Maximum likelihood APSL parameter estimates for different values of $\tau = 10^{-\varphi}$.

5.3.2.3 APSL Solution Uniqueness Analysis

In this subsection we briefly investigate the uniqueness of APSL choice probability solutions in the context of the Sioux Falls simulation study. Just as in Section 4.4, we plot trajectories of APSL solutions to approximate the uniqueness conditions, i.e. estimate β_{max}. A single simulation study is conducted for $\alpha_1^{true} = 0.3$, $\beta_{true} = 0.6$, and $Z = 2000$, leading to maximum likelihood estimates $\hat{\alpha}_1 = 0.3064$ and $\hat{\beta} = 0.6001$. We thus investigate whether APSL solutions are unique for these parameter estimates. Fig. 21 displays the maximum choice probability from three trajectories of APSL solutions as the β parameter is varied for four different randomly chosen OD movements, with $\alpha_1 = \hat{\alpha}_1 = 0.3064$. β was decremented by 0.01, and the initial large β value was $\beta = 2$. As shown, the $\beta_{max,m}$ values (β_{max} for OD movement m) for these OD movements can be estimated to vary between 0.86 and 0.94, suggesting that $\beta = 0.6001$ results in universally unique solutions.

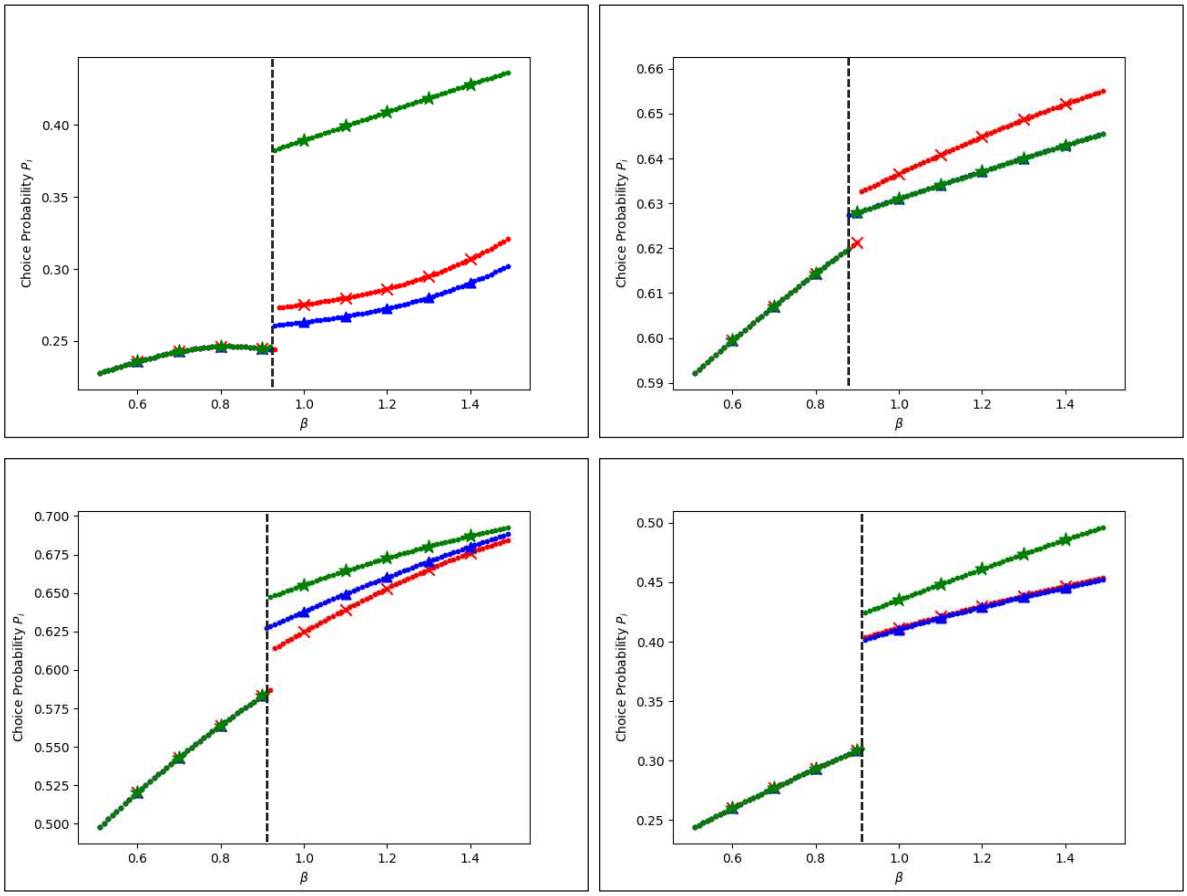


Fig. 21. Sioux Falls simulation study: Maximum choice probability of trajectories of APSL solutions as β is varied; $\alpha_1 = 0.3064$.

5.4 Real-Life Case Study

In this section we estimate the APSL model, where the model parameters are estimated using MLE with observed route choices tracked by GPS units. The data has been collected among drivers in the eastern part of Denmark in 2011, and includes a total of 17,115 observed routes. The dataset is the same as used in Prato et al (2014) as well as Rasmussen et al (2017), and after a filtering to include only trips where the sum of travel time (in minutes) and length (in km) is at least 10, a total of 8,696 observations remain.

The GPS traces are map matched to a network, for which corresponding time-of-day dependent travel times are available on the entire network. See more details in Prato et al (2014). The network is large-scale, representing all of Denmark, and thus includes 34,251 links. With current alternative generation techniques, it is not feasible to enumerate the universal choice set for such a large network. Instead, we approximate the universal choice set by generating a choice set for each observed route by applying the doubly stochastic approach also applied in Prato et al (2014). This approach is based on repeated shortest path search in which the network attributes and parameters of the cost function is perturbated between searches (Nielsen, 2000; Bovy & Fiorenzo-Catalano, 2007). Up to 100 unique paths are generated for each observation, see the distribution of number of alternatives in Fig. 22.

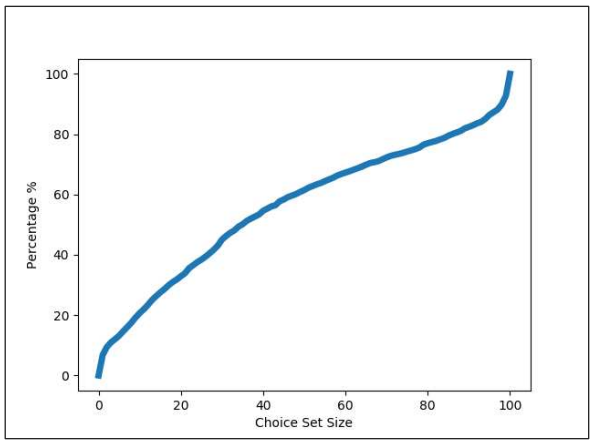


Fig. 22. Real-life case-study: Cumulative distribution of the choice set sizes for the 8,696 observations.

For the estimation, the travel cost of link a is specified as a weighted sum of congested travel time $w_{a,1}$ (in minutes), and length $w_{a,2}$ (in kilometres), such that:

$$t_a(\boldsymbol{w}_a; \boldsymbol{\alpha}) = w_{a,1} \cdot \alpha_1 + w_{a,2} \cdot \alpha_2$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$ are the congested travel time, and length parameters, respectively. The generalised travel cost for route $i \in R_m$ is thus:

$$c_{m,i}\big(\boldsymbol{t}(\boldsymbol{w}; \boldsymbol{\alpha})\big) = \sum_{a \in A_{m,i}} t_a(\boldsymbol{w}_a; \boldsymbol{\alpha}) = \sum_{a \in A_{m,i}} \big(w_{a,1} \cdot \alpha_1 + w_{a,2} \cdot \alpha_2\big) = \alpha_1 \sum_{a \in A_{m,i}} w_{a,1} + \alpha_2 \sum_{a \in A_{m,i}} w_{a,2}.$$

The model requires the specification of four parameters: α_1, α_2, θ, and β, but to ensure identification, the θ parameter is fixed at $\theta = 1$. Fig. 23A shows the relative travel time deviations away from the quickest routes in the choice sets for the observed routes as well as the alternative routes generated, and Fig. 23B shows the relative length deviations. 47% and 36% of the observed routes were the quickest and shortest routes, respectively. Moreover, there appear to be observations of unattractive route choices, where some observed routes were 2.11 times slower / 2.29 times longer than the quickest / shortest alternatives, as well as numerous potentially unrealistic routes generated, where some generated routes are 2.91 times slower / 3.22 times longer.

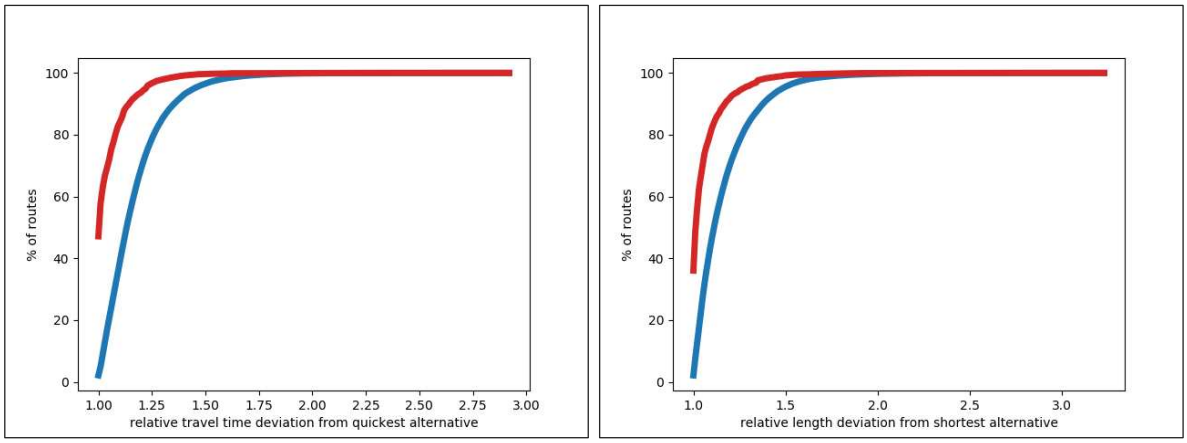


Fig. 23. Real-life case-study: Relative deviations away from quickest/shortest routes in the choice sets for the observed routes (red) and alternative routes generated (blue) – A: Travel time. B: Length.

We estimate the models utilising the same Log-Likelihood maximisation algorithm (L-BFGS-B, see Section 5.2.2), initial conditions, and parameter bounds, where appropriate. Initial conditions: $\left(\tilde{\alpha}_1^{(1)}, \tilde{\alpha}_2^{(1)}, \tilde{\beta}^{(1)}, \tilde{\lambda}^{(1)}\right) = (0.5,0.5,0,0)$, and bounds: $\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta} \in [0,2]$, $\tilde{\lambda} \in [0,200]$.

5.4.1 APSL Estimation

5.4.1.1 Results

In this subsection we provide results from estimating the three parameters of the APSL model in this case study: α_1, α_2, and β. Table 3 displays the APSL parameter estimates and the consequent Log-Likelihood value.

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}$	LL
0.633	0.184	0.840	-18978

Table 3. Real-life case-study: APSL parameter estimates and Log-Likelihood.

Fig. 24 shows the Log-Likelihood surfaces around the three parameter estimates; as can be seen, these are smooth and maximal around the estimates.

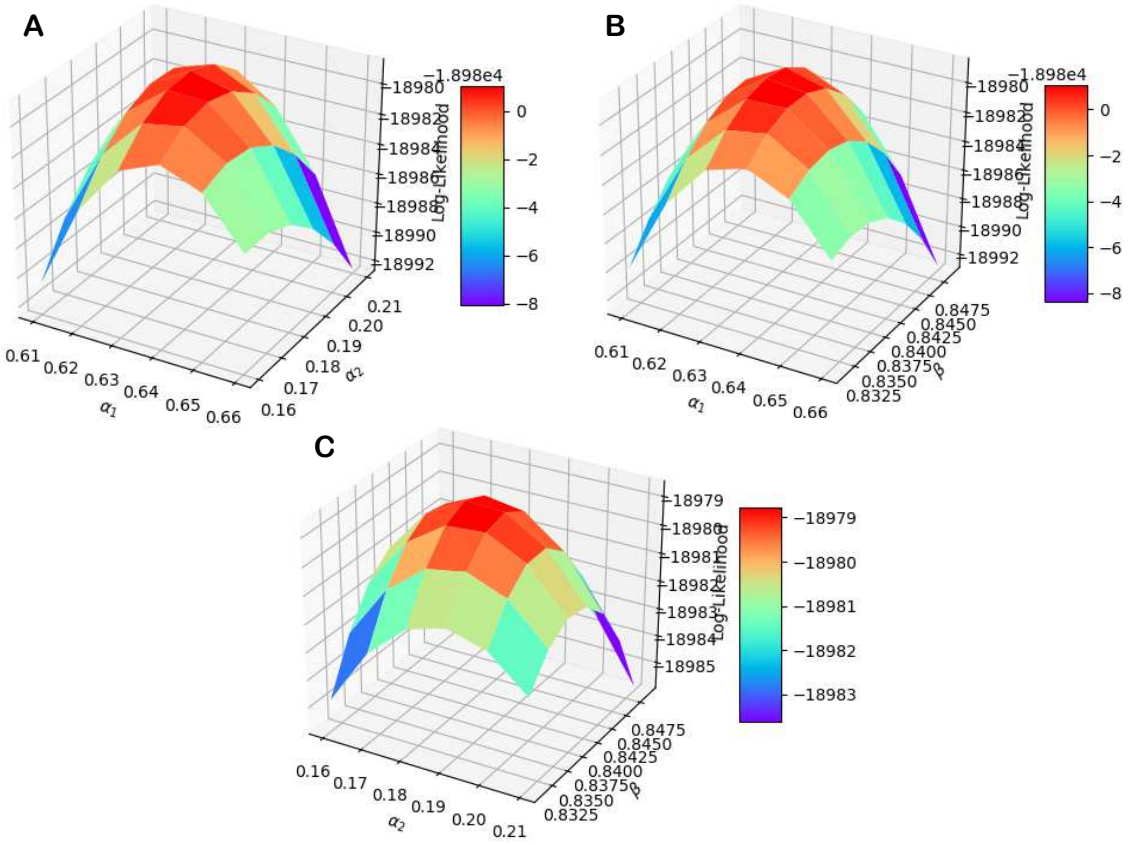


Fig. 24. Real-life case-study: APSL Log-Likelihood surfaces around parameter estimates in Table 3 – A: α_1, α_2. B: α_1, β. C: α_2, β.

5.4.1.2 Computation Analysis

We analyse here the computational performance of the APSL model in the real-life case study. The same computer was used as in Section 5.3.2.2.

Fig. 25A shows for different values of the APSL choice probability convergence parameter ξ (and thus convergence statistic), the average number of fixed-point iterations per OD movement and computation time required to solve all of the 8,696 APSL fixed-point problems $\boldsymbol{P}_{m_z} = \boldsymbol{G}_{m_z}\left(\boldsymbol{g}_{m_z}\left(\boldsymbol{c}_{m_z}(\boldsymbol{t}), \boldsymbol{\gamma}_{m_z}^{APS}(\boldsymbol{t}, \boldsymbol{P}_{m_z})\right)\right)$ for $z = 1, \ldots, Z$, and consequently compute a single Log-Likelihood, with the estimated APSL parameters in Table 3. Fig. 25B shows the value of the Log-Likelihood obtained as ξ is increased. As shown, computation time and average number of fixed-point iterations per OD increase linearly as the convergence parameter is increased, and the Log-Likelihood increases in accuracy (from $\xi = 2$) as the APSL choice probabilities become more accurate. The relatively large estimated β value results in a longer computation time, as shown in Section 5.3.2.2, for lower β, the computation times are less.

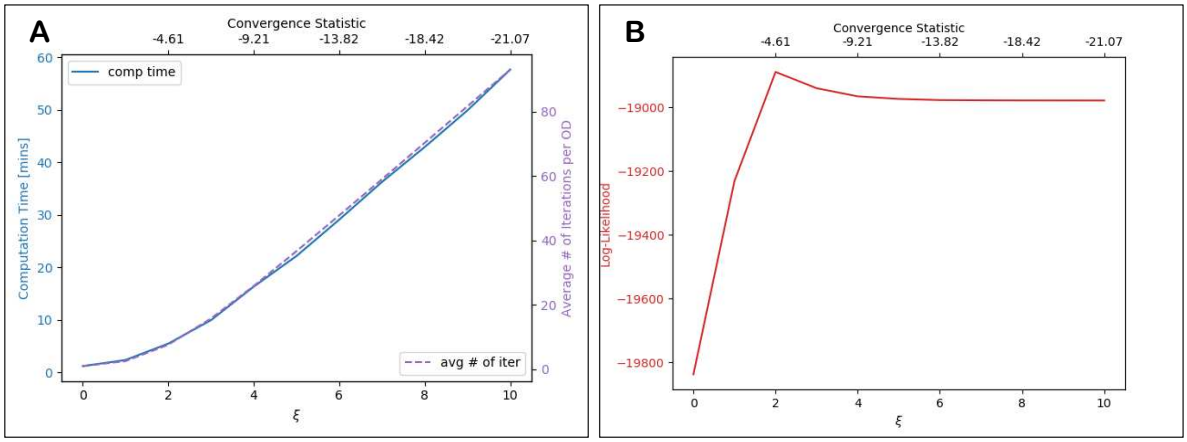


Fig. 25. Real-life case-study: Computational statistics for calculating APSL Log-Likelihoods as the APSL choice probability convergence parameter ξ is increased –

A: Average number of fixed-point iterations per OD / computation time [mins]. B: Log-Likelihood value.

Fig. 26A-B show for a single estimation of the APSL model (implementation of the L-BFGS-B algorithm), the cumulative computation times of the iterations and the Log-Likelihood values and parameter estimates at the end of each iteration, respectively, with $\xi = 10$. The initial conditions for solving the APSL fixed-point problems were updated at the end of each iteration with the choice probabilities obtained from the current parameter estimates.

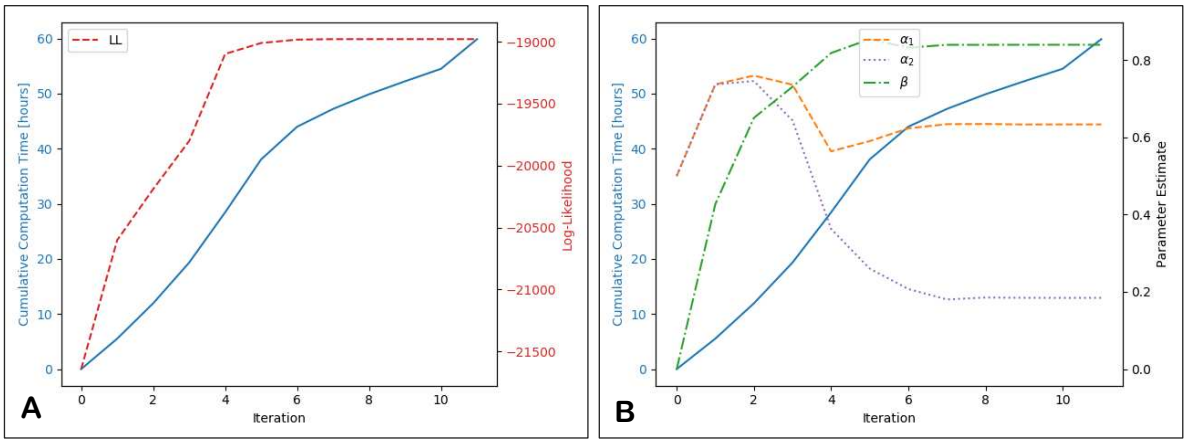
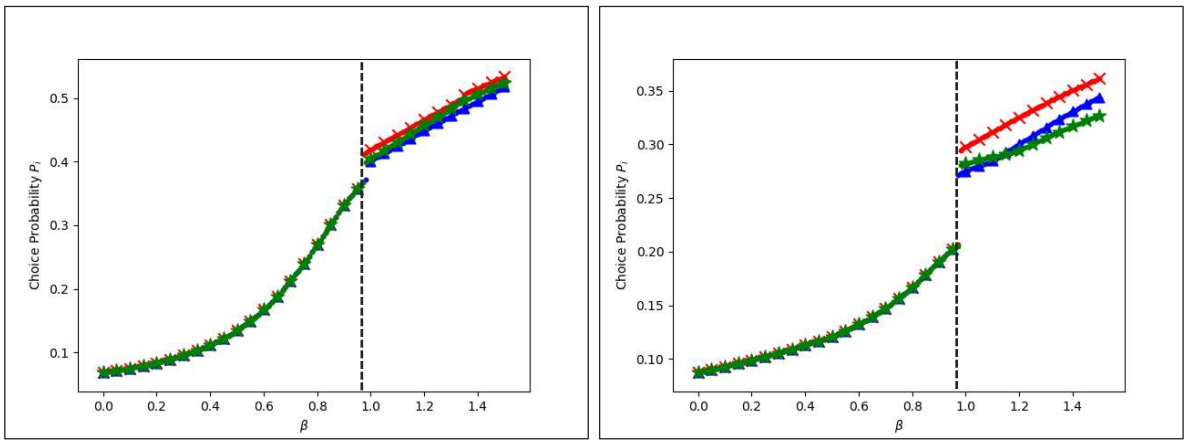


Fig. 26. Real-life case-study: Cumulative computation time at each iteration for a single estimation of the APSL model, and MLE statistics; $\xi = 10$ – A: Log-Likelihood. B: Parameter estimates.

5.4.1.3 APSL Solution Uniqueness Analysis

We briefly investigate here the uniqueness of APSL choice probability solutions in the context of the real-life case study. Similar to the experiments conducted in Section 5.3.2.3 for the Sioux Falls simulation study, we estimate the uniqueness conditions for the network given the estimated parameters. Trajectories of APSL solutions are plotted to approximate β_{max}. Fig. 27 displays the maximum choice probability from trajectories of APSL solutions as the β parameter is varied for four different randomly chosen OD movements, with α_1 and α_2 as in Table 3. β was decremented by 0.01, and the initial large β value was $\beta = 1.5$. As shown, the $\beta_{max,m}$ values (β_{max} for OD movement m) for these OD movements are between 0.96 and 0.99, suggesting that $\beta = 0.84023$ results in universally unique solutions.



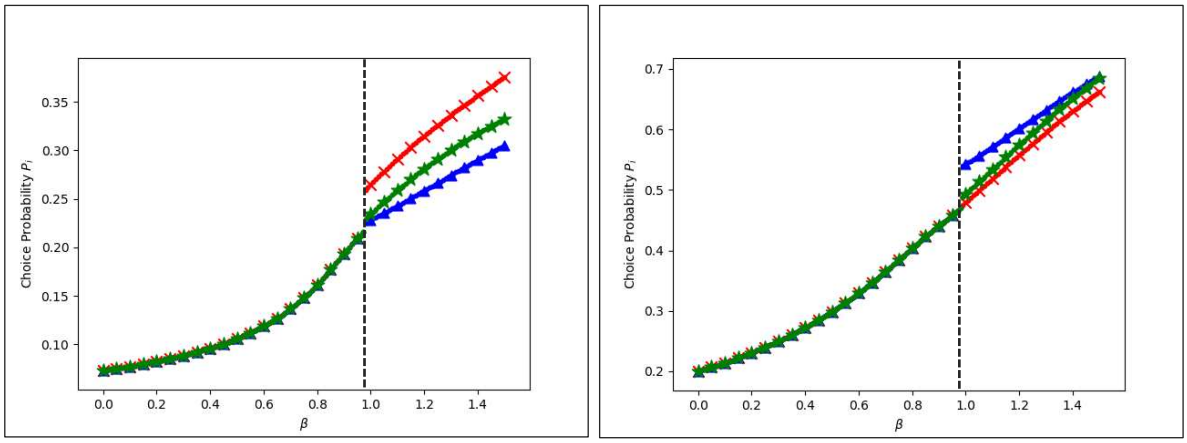


Fig. 27. Real-life case study: Maximum choice probability from trajectories of APSL solutions as β is varied; $\alpha_1 = 0.63333$, $\alpha_2 = 0.18428$.

5.4.2 Comparing Results with Other Path Size Logit Models

In this subsection we estimate models discussed in this paper and compare results. Table 4 shows the estimated parameters for the MNL, PSL, GPSL, GPSL$'_{(\lambda=\theta)}$, and APSL models.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}$	$\hat{\lambda}$	LL
MNL	0.777	0.330			-21308
PSL	0.966	0.306	1.347		-20581
GPSL	0.415	0.085	1.186	91.95	-17874
GPSL$'_{(\lambda=\theta)}$	0.691	0.154	1.807		-19152
APSL	0.633	0.184	0.840		-18978

Table 4. Real-life case-study: Estimation results and stability statistics from all Path Size Logit models.

To compare the estimation results of the models, we apply the approach in Swait & Ben-Akiva (1984) based on the non-nested test in Horowitz (1983) in combination with the results in Table 4. The adjusted rho-squared for model h with estimated parameters $\hat{\boldsymbol{\omega}}_h$ is given by:

$$\bar{\rho}_h^2 = 1 - \frac{LL_h(\hat{\boldsymbol{\omega}}_h) - K_h}{LL^*},$$

where $LL_h(\hat{\boldsymbol{\omega}})$ is the Log-Likelihood for model h given the estimated parameters $\hat{\boldsymbol{\omega}}_h$, K_h is the number of model h parameters, and LL^* is the equal choice probability Log-Likelihood which in this case study is: $LL^* = \ln\left(\prod_{z=1}^{Z} \frac{1}{N_{m_z}}\right) = -\sum_{z=1}^{Z} \ln(N_{m_z}) = -28200$. The distribution of the difference between $\bar{\rho}_{h_1}^2$ and $\bar{\rho}_{h_2}^2$ for models h_1 and h_2, respectively, (which are possibly non-nested) is given by:

$$\Pr\left(\bar{\rho}_{h_2}^2 - \bar{\rho}_{h_1}^2 > y\right) \leq \Phi\left(-\left[-2yLL^* + \left(K_{h_2} - K_{h_1}\right)\right]^{\frac{1}{2}}\right),$$

where $y > 0$ is the test statistic. To test the null hypothesis that the MNL model outperforms the PSL, GPSL, GPSL$'_{(\lambda=\theta)}$, and APSL models, we compute the test statistics $y_{PS} = \bar{\rho}_{PS}^2 - \bar{\rho}_{MNL}^2$, $y_{GPS} = \bar{\rho}_{GPS}^2 - \bar{\rho}_{MNL}^2$, $y_{GPS'} = \bar{\rho}_{GPS'}^2 - \bar{\rho}_{MNL}^2$, and $y_{APS} = \bar{\rho}_{APS}^2 - \bar{\rho}_{MNL}^2$. Similarly, we compute the corresponding $Y_{PS} = -[-2y_{PS}LL^* + (K_{PS} - K_{MNL})]^{\frac{1}{2}}$, $Y_{GPS} = -[-2y_{GPS}LL^* + (K_{GPS} - K_{MNL})]^{\frac{1}{2}}$, $Y_{GPS'} = -[-2y_{GPS'}LL^* + (K_{GPS'} - K_{MNL})]^{\frac{1}{2}}$, and $Y_{APS} = -[-2y_{APS}LL^* + (K_{APS} - K_{MNL})]^{\frac{1}{2}}$. The results are shown in Table 5. $\Pr(y_h \leq Y_h)$ is the probability that the MNL model outperforms model h, but these values are too small for computer precision to calculate. This exemplifies the necessity of capturing the correlation between routes. One can identify however that $\Pr(y_{PS} \leq Y_{PS}) > \Pr(y_{GPS'} \leq Y_{GPS'}) > \Pr(y_{APS} \leq Y_{APS}) > \Pr(y_{GPS} \leq Y_{GPS})$. In another comparison of fit test, Table 6 shows the penalised-likelihood criteria.

In both tests, the GPSL$'_{(\lambda=\theta)}$ and APSL models outperform the PSL model with the same number of parameters, where APSL outperforms GPSL$'_{(\lambda=\theta)}$. This suggests that there is value in including a measure of distinctiveness within the path size contribution factors. The GPSL model outperforms all models due to the greater flexibility the λ parameter provides. Several case studies have found that larger values of λ increase the goodness-of-fit of the GPSL model

(Ramming, 2002; Prato, 2005; Hoogendoorn-Lanser, 2005; Bekhor & Prato, 2006), and hence it is not unusual that $\hat{\lambda} = 91.95$ is so big. We explore further below.

h	K_h	$\bar{\rho}_h^2$	y_h	Y_h
PSL	3	0.27006	0.02573	-38.1144
GPSL	4	0.36604	0.12171	-82.8673
$\text{GPSL}'_{(\lambda=\theta)}$	3	0.32074	0.07642	-65.6583
APSL	3	0.32690	0.08258	-68.2537

Table 5. Real-life case-study: Comparison of fit between models based on non-nested Horowitz type tests.

	AIC	BIC	CAIC
MNL	42621	42635	42637
PSL	41169	41190	41193
GPSL	35756	35784	35788
$\text{GPSL}'_{(\lambda=\theta)}$	38311	38332	38335
APSL	37963	37984	37987

Table 6. Real-life case-study: Comparison of fit between models based on penalised-likelihood criteria.

For very large values of λ within the GPSL path size terms, the path size contributions become extremely sensitive to differences in cost, where the contribution of route k to the path size term of route i is large if $c_i > c_k$ and small if $c_i < c_k$, and as $\lambda \to \infty$, $\left(\frac{c_i}{c_k}\right)^{\lambda} \to \infty$ if $c_i > c_k$, and $\left(\frac{c_i}{c_k}\right)^{\lambda} \to 0$ if $c_i < c_k$. The implication of this is that routes with relatively small travel costs are penalised significantly less than routes with relatively large travel costs for link sharing, and hence that low costing routes are considered much more distinct than high costing routes.

To provide some measure of the relative cost and distinctiveness of the observed routes (in comparison with the generated alternatives), Fig. 28A plots the percentage of generated routes with a travel cost greater than the observed route in each choice set (where the GPSL travel cost parameters are used), against the percentage of routes with PSL path size terms smaller than the observed route. The PSL path size terms provide a measure of the universal distinctiveness of the alternatives, i.e. without considering whether or not the routes are link sharing with unrealistic alternatives. The bottom right of the figure appears to be highly populated suggesting that many of the observations have relatively low travel costs but are relatively universally indistinct, while a sizeable proportion are relatively distinct, even without the contribution weighting. Most notably though, a considerable proportion of the route observations have a low percentage of routes with greater travel costs, and many of these are relatively universally distinct (top left of figure). This perhaps suggests that many drivers have taken unattractive, relatively costly routes that are distinct. Fig. 28B plots the same cost percentage against the GPSL path size term percentage. As expected, the route observations with low costs are now considered much more distinct, while the observations with large costs are considered less distinct.

Fig. 29 displays the choice probability distribution of the observed routes under the different estimated models. For all models, a large percentage of the observed routes have small choice probabilities. This seems to also suggest that there are many observations where an unattractive route was chosen.

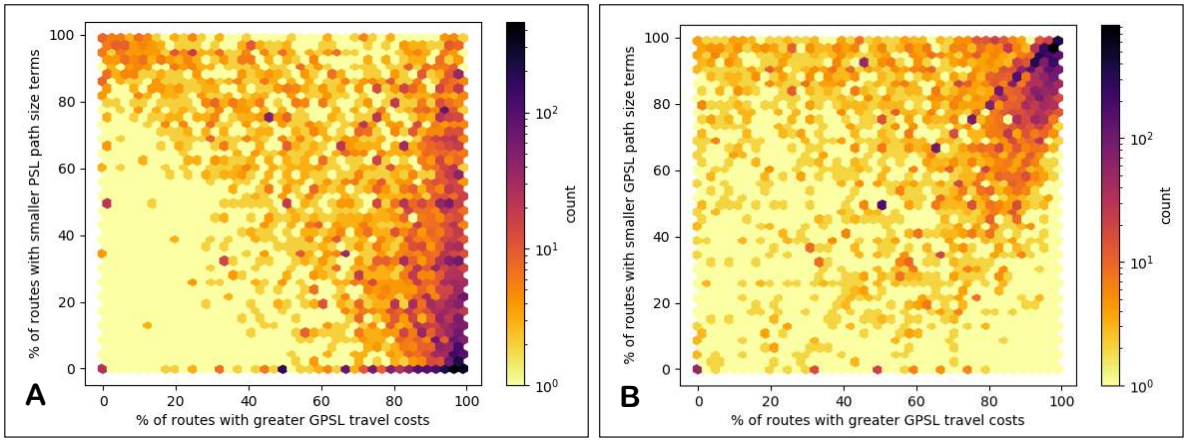


Fig. 28. Real-life case study: Percentage of routes in each OD movement choice set with costs greater / path size terms smaller than the observed route – A: PSL path size terms. B: GPSL path size terms.

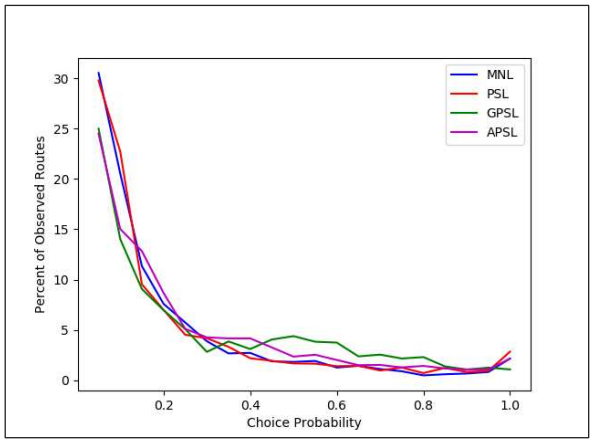
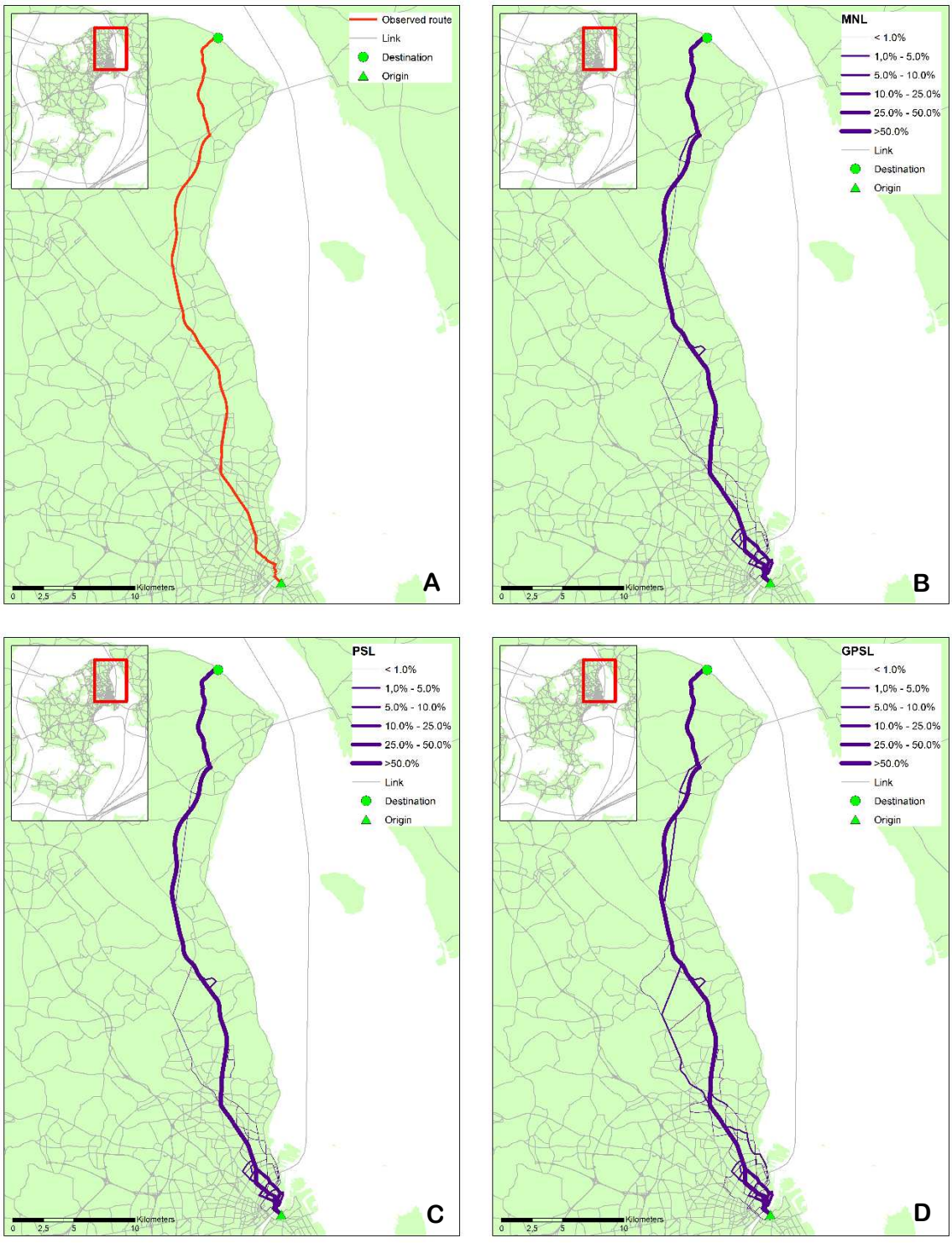


Fig. 29. Real-life case study: Choice probability distribution of the observed routes under the different estimated models.

The data set contains relatively costly but relatively universally distinct route observations. The GPSL model is able to provide the best fit for these observations, without compromising the fit for the low costing observations. The GPSL travel cost parameter estimates are smaller than the same estimates for the other models, which improves the relative attractiveness of the costly alternatives. To counterbalance this so that that the low costing routes still remain attractive, GPSL introduces a large λ value: routes with relatively small travel costs are penalised significantly less than routes with relatively large travel costs for link sharing. Moreover, GPSL is able to further increase the relative attractiveness of the distinct, costly routes by decreasing the attractiveness of the indistinct, costly routes with the large λ.

Fig. 30A & Fig. 31A show two route observations, which we label OD 1 and OD 2, respectively. Fig. 30B-D & Fig. 31B-D plot the consequent link choice probabilities from the MNL, PSL, GPSL, and APSL models. From first inspection it appears that the route taken by the driver in OD 1 is a high probability, attractive route, while the route taken in OD 2 is low probability. Table 7 displays the choice probabilities of the observed route for OD 1 and OD 2 under the different models. The APSL model provides the largest choice probability for the observed route in OD 1, and GPSL provides the highest for OD 2, where the chosen probabilities for OD 2 are small.


Observed route
Link
Destination
Origin
A
MNL
< 1.0%
1,0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
B
PSL
< 1.0%
1,0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
C
GPSL
< 1.0%
1,0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
D
Kilometers
0 2,5 5 10

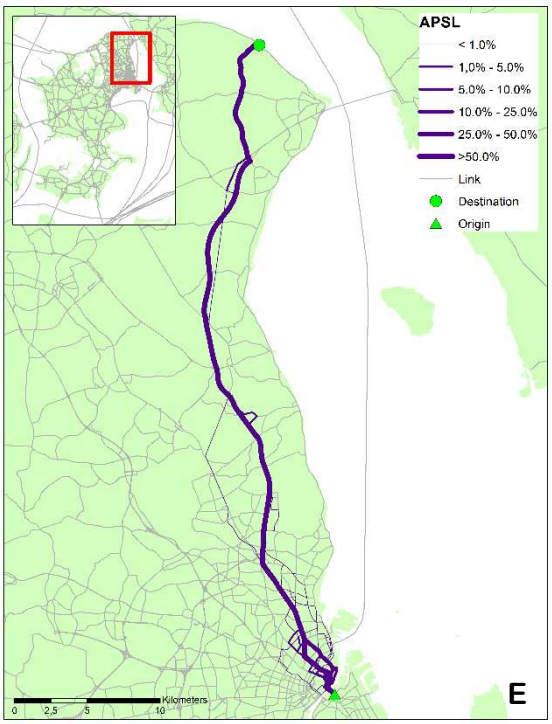
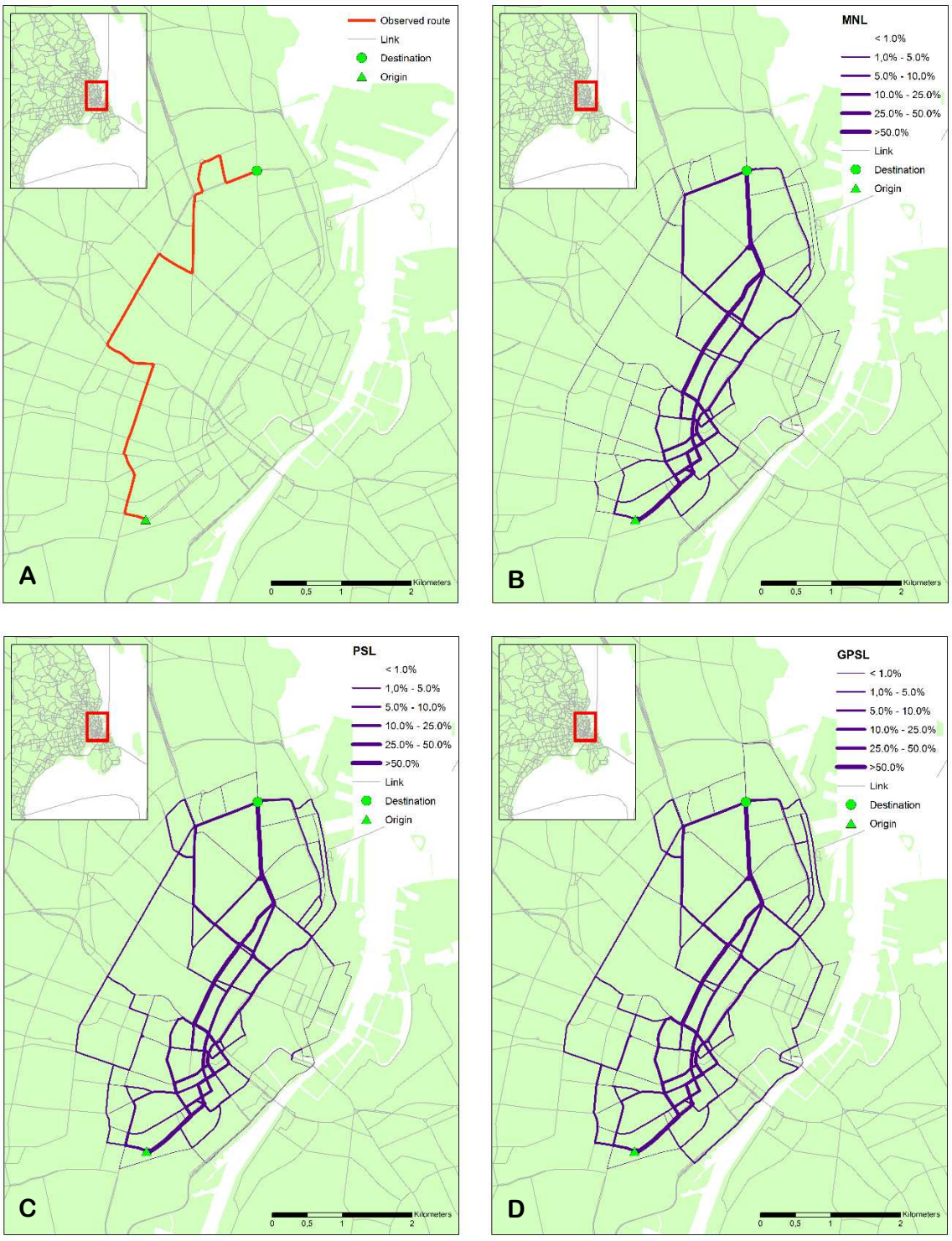


Fig. 30. Real-life case study: OD 1 plotted link choice probabilities from the estimated models for a single observation – A: Observed route. B: MNL. C: PSL. D: GPSL. E: APSL.


Observed route
Link
Destination
Origin
A
MNL
< 1.0%
1.0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
B
PSL
< 1.0%
1.0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
C
GPSL
< 1.0%
1.0% - 5.0%
5.0% - 10.0%
10.0% - 25.0%
25.0% - 50.0%
>50.0%
Link
Destination
Origin
D
Kilometers
0 0,5 1 2

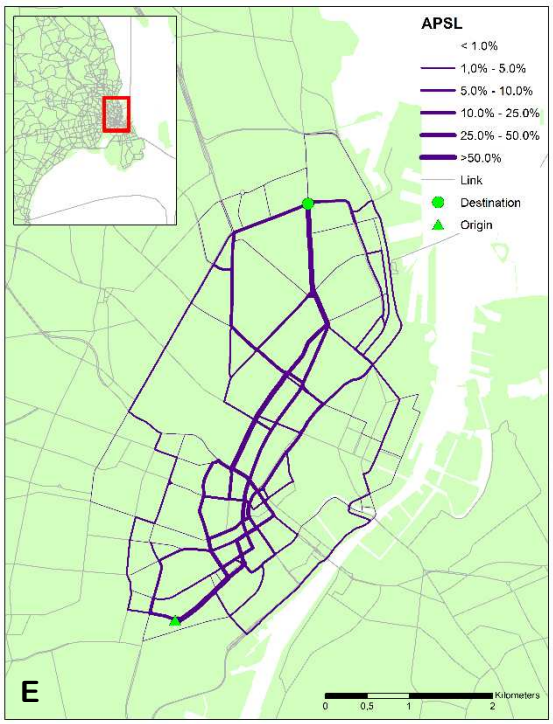


Fig. 31. Real-life case study: OD 2 plotted link choice probabilities from the estimated models for a single observation – A: Observed route. B: MNL. C: PSL. D: GPSL. E: APSL.

	MNL	PSL	GPSL	APSL
OD 1	0.101	0.110	0.150	0.261
OD 2	0.0004	0.0024	0.0042	0.0025

Table 7. Real-life case study: OD 1 & OD 2 observed route choice probabilities for the different estimated models.

For Path Size Logit models, the utility for route $i \in R_m$ is comprised of a cost component $-\theta c_{m,i}$ and a path-size component $\beta \ln(\gamma_{m,i})$, i.e. so that the utility is $V_{m,i} = -\theta c_i + \beta \ln(\gamma_i)$. Fig. 32A-B plot for OD 1 and OD 2, respectively, the cost components against path size components of the routes under the PSL, GPSL, and APSL models, where the observed route is in red. Fig. 33A-B plot the cost and path size components against choice probability. In both cases, universal distinctiveness tends to increase as travel cost increases. For OD 1, the observed route is universally indistinct but low costing, and is the highest choice probability route for all models. The APSL model thus provides the best fit for this observation: the path size contribution factors consider probability ratios and hence the observed route is considered the most attractive and distinct compared to its overlapping routes. The GPSL model reduces the range for the cost components and hence decreases the attractiveness of the observed route according to its cost, which is not compensated for by its distinctiveness, since it is highly correlated with other low costing routes. For OD 2, the observed route is universally distinct but high costing, and has a low choice probability for all models. As discussed above, the GPSL model is able to provide the best fit for these observations.

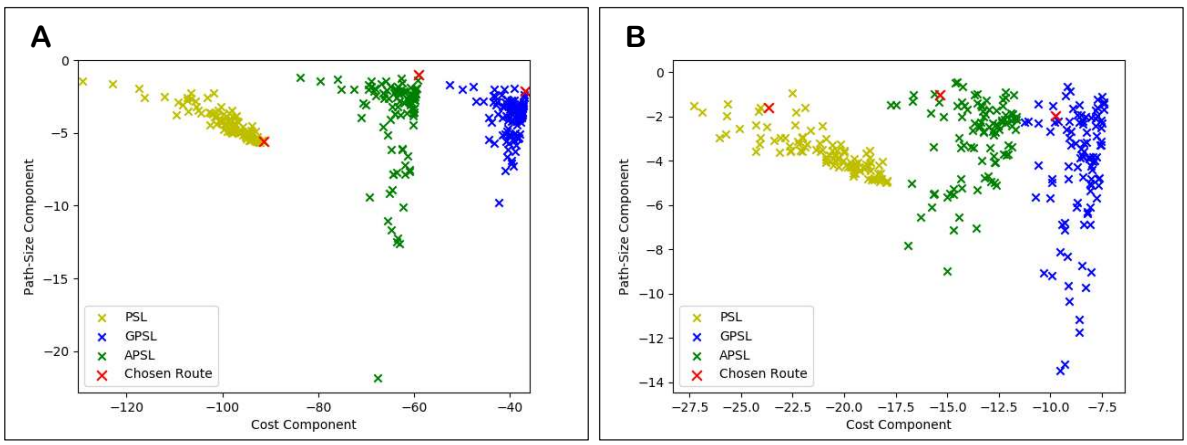


Fig. 32. Real-life case study: Cost / Path-Size components of the route utilities from PSL/GPSL/APSL – A: OD 1. B: OD 2.

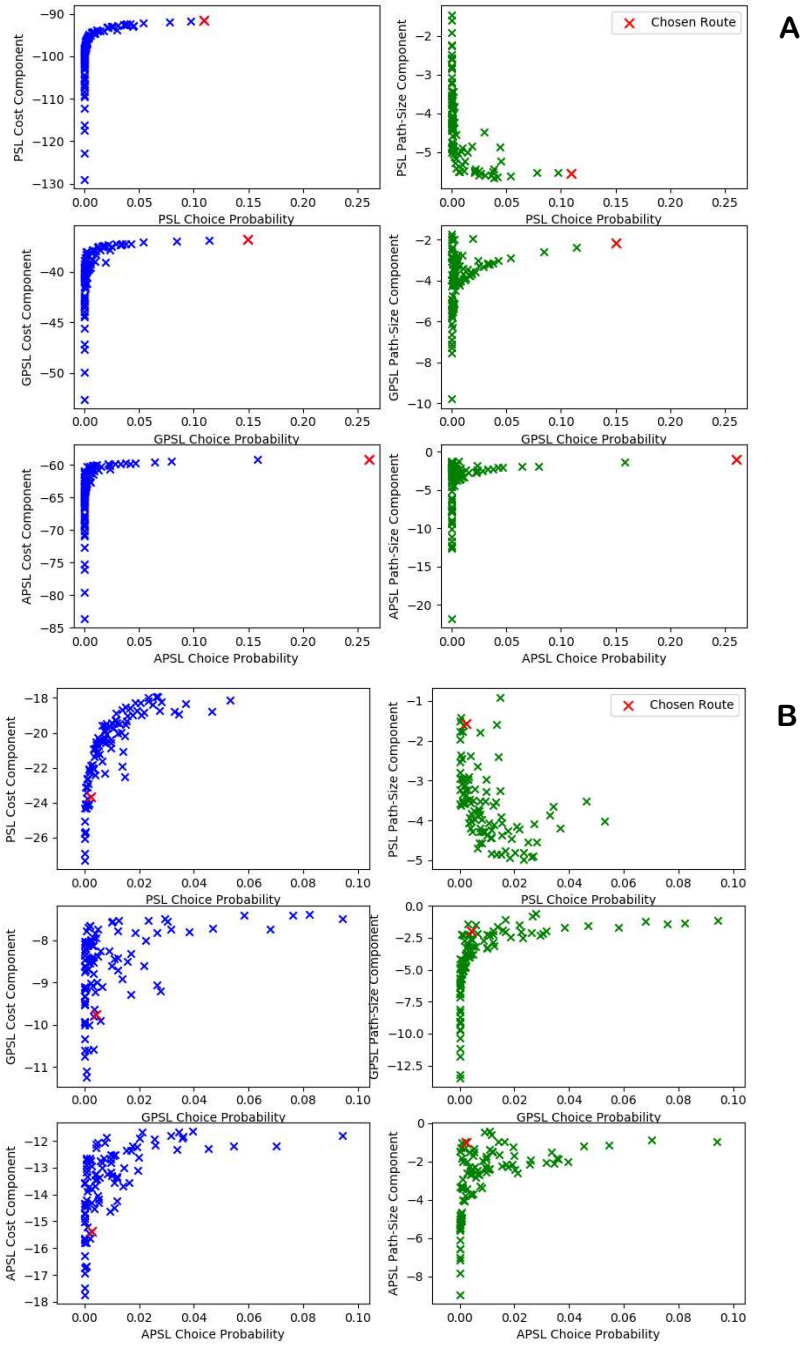


Fig. 33. Real-life case study: Choice probability against cost / path-size component for PSL/GPSL/APSL – A: OD 1. OD 2.

The consistency requirement we impose on the APSL model constrains the way that it can mimic the behaviour of the GPSL model, since if parameters are chosen to capture the high costing observations, a price is paid in terms of the feedback effect to the path size correction terms. This confirms that APSL and GPSL are quite different candidate models in the way in which they aim to capture behaviour.

The λ parameter allows the GPSL model to improve the choice probabilities of high costing, distinct route observations one might consider as being outliers / route choice decisions made according to unobserved attributes, though not by design. It seems unlikely that the GPSL model was formulated anticipating extremely large values of λ (such as 91.95) given the exponential nature of the path size contribution factors. In fact, Ramming (2002) estimates the proposed exponential formulation and finds $\lambda = \infty$ provides the best fit to the data, which does not seem reasonable. Moreover, Ramming (2002) hypothesises that the path size contribution factor should 'split the link size contributions more severely than MNL would split path shares, or counter-intuitive predictions will result', though a reason is not

given. It's difficult to know what a 'sensible' restriction would be that one could impose upon the λ parameter so that the GPSL model behaves according to a more feasible theoretical interpretation. Bekhor & Prato (2006) utilise $\lambda = 9$, Hoogendoorn-Lanser (2005) utilises $\lambda = 20$, and Prato (2009) claims optimal values vary between 10 and 15, though these values still seem large. Table 8 displays the results from estimating the GPSL model with a restriction imposed upon λ so that $\lambda \leq 10$. As anticipated, the optimal value for λ is at the bound (equal to 10), but the Log-Likelihood value no longer beats the $\text{GPSL}'_{(\lambda=\theta)}$ and APSL models. What is also interesting is that the estimated β value is almost 2, which is very large seeming the theory suggests it should be around 1. This supports the theory that the GPSL path size components are capturing something other than the correlation.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}$	$\hat{\lambda}$	LL
MNL	0.777	0.330			-21308
PSL	0.966	0.306	1.347		-20581
GPSL ($\lambda \leq 10$)	0.769	0.160	1.943	10	-19312
GPSL	0.415	0.085	1.186	91.95	-17874
$\text{GPSL}'_{(\lambda=\theta)}$	0.691	0.154	1.807		-19152
APSL	0.633	0.184	0.840		-18978

Table 8. Real-life case-study: Estimation results including GPSL with λ restricted to $\lambda \leq 10$.

6 Summary and Scope for Further Research

Due to their comparatively low computational cost and relative ease in obtaining reasonable estimates for parameters, Path Size Logit route choice models are a useful and practical approach to approximating the correlation between routes. Existing Path Size Logit models, however, have some key theoretical weaknesses: for PSL the presence of unrealistic routes in a choice set negatively impacts the choice probabilities of realistic routes when links are shared, and for GPSL there are internal inconsistency issues which can have negative implications, for example routes which are defined as unrealistic by the path size terms may not be routes with low choice probabilities. The intricacies of the issues with existing Path Size Logit models are demonstrated in the paper, and a new APSL model is proposed which provides a potential solution to these issues. The APSL model proposes that routes contribute to path size terms according to probability ratios, and choice probability solutions to the model are solutions to the fixed-point problem involving the probabilities.

The paper proves that choice probability solutions to the APSL model are guaranteed to exist, and proves that values of b exist such that APSL solutions are unique for β in the range $0 \leq \beta \leq b$. Though there are cases where solutions are unique for all $\beta \geq 0$, in most cases there is a maximum value for b (b_{max}). β in the range $0 \leq \beta \leq b_{max}$ is however only a sufficient condition for unique APSL solutions, β_{max} is the true maximum value where solutions are unique for β in the range $0 \leq \beta \leq \beta_{max}$, and a method is proposed in the paper for estimating β_{max}.

To show that the parameters of the APSL model can be estimated, a Maximum Likelihood Estimation procedure is proposed for estimating APSL with tracked route observation data. This procedure is then first investigated in a simulation study on the Sioux Falls network where it is shown that it is generally possible to reproduce assumed true parameters. The APSL model is then estimated using real tracked route GPS data on a large-scale network. Results show that the APSL outperforms the MNL and PSL models with the same number of model parameters, while the GPSL model outperforms APSL due to the added flexibility an additional parameter provides.

The APSL model requires a fixed-point algorithm to approximate solutions. The paper assesses the computational performance of the FPIM for calculating choice probabilities and estimating the parameters of the APSL model, where accuracy is compared with computation time. Results indicate that accurate choice probability solutions and parameter estimates can be obtained from feasible computation times.

Future research should explore the application of the APSL model within a Stochastic User Equilibrium framework, which could involve exploring whether one can combine the fixed-point iterations used for APSL with those used for congestion, so that they are performed simultaneously.

As noted in our numerical experiments, the consistency condition that we impose in the APSL model, while offering improvements over PSL, constrains the extent to which the model is able to compete with the GPSL model in terms of model-fit, with the additional parameter in the GPSL model allowing it to de-couple the scale of the model from the path-size effect, albeit at the price of inconsistency. A natural path for future research could be to explore the potential for developing generalised forms of APSL, in the spirit of GPSL, allowing an extra dimension (parameter) to fit, but without sacrificing the requirement for consistency.

7 Acknowledgements

We gratefully acknowledge funding provided by the University of Leeds by awarding the corresponding author with a University of Leeds Doctoral Scholarship for PhD research, and, the financial support of the Independent Research Fund

Denmark to the project “Using Big Data sources for the consistent estimation of next-generation route choice models”. We would also like to thank three anonymous referees for their highly constructive criticism, which helped in improving earlier versions of this paper.

8 References

Ahipasaoglu S, Meskarian R, Magnanti T.L, & Natarajan K, (2015). Beyond normality: A cross moment stochastic user equilibrium model. Transportation Research Part B, 81(2), p.333-354.

Ahipasaoglu S, Arikan U, Natarajan K, (2016). On the flexibility of using marginal distribution choice models in traffic equilibrium. Transportation Research Part, 91, p.130-158.

Bekhor S, & Prashker J, (1999). Formulations of extended logit stochastic user equilibrium assignments. In: Proceedings of the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, p.351–372.

Bekhor S, & Prashker J, (2001) Stochastic user equilibrium formulation for the generalized nested logit model. Transportation Research Record, 1752, p.84–90.

Bekhor S, & Prato C, (2006). Effects of choice set composition in route choice modeling. Proceedings of the 11th International Conference on Travel Behavior Research, Kyoto, Japan.

Bekhor S, Ben-Akiva M, & Ramming M, (2002). Adaptation of logit kernel to route choice situation. Transportation Research Record, 1805, p.78–85.

Bekhor S, Chorus C, & Toledo T, (2012). Stochastic user equilibrium for route choice model based on random regret minimization. Transportation Research Record, 2284(1), p.100-108.

Ben-Akiva M, & Ramming S, (1998). Lecture notes: discrete choice models of traveler behavior in networks. Prepared for Advanced Methods for Planning and Management of Transportation Networks. Capri, Italy.

Ben-Akiva M, & Bierlaire M, (1999). Discrete choice methods and their applications to short term travel decisions. In: Halled, R.W. (Ed.), Handbook of Transportation Science. Kluwer Publishers.

Ben-Akiva M, & Bolduc D, (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper.

Bierlaire M, (2002). The Network GEV Model. Conference paper STRC 2002.

Bovy P, & Fiorenzo-Catalano S, (2007). Stochastic Route Choice Set Generation: Behavioral and Probabilistic Foundations. Transportmetrica, 3, p.173–189.

Bovy P, Bekhor S, & Prato C, (2008). The Factor of Revisited Path Size: Alternative Derivation. Transportation Research Record: Journal of the Transportation Research Board, 2076, Transportation Research Board of the National Academies, Washington, D.C., p.132–140.

Byrd R, Lu P, Nocedal J, & Zhu C (1994). A Limited Memory Algorithm for Bound Constrained Optimization. Technical Report NAM-08, Northwestern University, Department of Electrical Engineering and Computer Science.

Cantarella G, & Binetti M, (2002). Stochastic assignment with gammit path choice models. Patriksson, M., Labbé’s, M. (Eds.), Transportation Planning: State of the Art, p.53–68.

Cascetta E, Nuzzolo A, Russo F, & Vitetta A, (1996). A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Leon, France, p.697–711.

Castillo E, Menéndez J, Jiménez P, & Rivas A, (2008). Closed form expression for choice probabilities in the Weibull case. Transportation Research Part B, 42(4), p.373–380.

Chikaraishi M, & Nakayama S, (2016). Discrete choice models with q-product random utilities. Transportation Research Part B, 93, p.576–595.

Chorus C, (2010). A New Model of Random Regret Minimization. EJTIR, 10(2), p.181-196.

Chu C, (1989). A paired combinatorial logit model for travel demand analysis. In: Proc. Fifth World Conference on Transportation Research, Ventura, Calif. 4, p.295–309.

Connors R, Hess S, & Daly A, (2014). Analytic approximations for computing probit choice probabilities, Transportmetrica A: Transport Science, 10(2), p.119-139.

Daganzo C, & Sheffi Y, (1977). On stochastic models of traffic assignment. Transportation Science, 11, p.253–274.

Daly A, & Bierlaire M, (2006). A general and operational representation of Generalised Extreme Value models. Transportation Research Part B, 40, p.285–305.

Damberg O, Lundgren J, & Patriksson M, (1996). An algorithm for the stochastic user equilibrium problem, Transportation Research, 30B, p.115–131.

Dial R, (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research, 5(2), p.83-111.

Frejinger E, & Bierlaire M, (2007). Capturing correlation with subnetworks in route choice models. Transportation Research Part B, 41(3), p.363-378.

Fosgerau M, & Bielaire M, (2009). Discrete choice models with multiplicative error terms. Transportation Research Part B, 43, p.494-505.

Gliebe J, Koppleman F, & Ziliaskopoulos A, (1999). Route choice using a paired combinatorial logit model. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, DC.

Horowitz J, (1983). Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. Transportation Science, 17(3),p.319-350

Hoogendoorn-Lanser S, (2005). Modelling travel behaviour in multi-modal networks. Ph.D. Thesis, TRAIL Research School, Technical University of Delft, The Netherlands.

Isaacson E, & Keller H, (1966). Analysis of Numerical Methods. John Wiley & Sons, Inc., New York, USA.

Johnson L, & Scholz D, (1968). On Steffensen's Method. SIAM Journal on Numerical Analysis, 5 (2), p.296-302.

Kitthamkesorn S, & Chen A, (2013). Path-size weibit stochastic user equilibrium model. Transportation Research Part B, 57, p.378-397.

Kitthamkesorn S, & Chen A, (2014). Unconstrained weibit stochastic user equilibrium with extensions. Transportation Research Part B, 59, p.1-21.

Li B, (2011). The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. Transportation Research Part B, 45, p.461–473.

Manzo S, Prato C & Nielsen O, (2015). How uncertainty in input and parameters influences transport model outputs: a four-stage model case-study. Elsevier, Transport Policy, 38, p.64-72.

Marzano V & Papola A (2008). On the covariance structure of the cross-nested logit model. Transportation Research Part B, 42(2), p.83–98

McFadden D, & Train K, (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics, 15 (5), p.447–470.

McFadden D, (1978). Modeling the choice of residential location. In: Anders, Karlqvist., Lars, Lundqvist., Folke, Snickars., Jorgen, Weibull. (Eds.), Special interaction theory and planning models. North-Holland, Amsterdam, p.75–96.

Nakayama S, & Chikaraishi M, (2015). Unified closed-form expression of logit and weibit and its extension to a transportation network equilibrium assignment. Transportation Research Part B, 81, p.672–685.

Natarajan K, Song M, & Teo C.P, (2009). Persistency Model and Its Applications in Choice Modeling. Management Science, 55(3), p.453-469

Nielsen O, & Frederiksen R, (2006). Optimisation of timetable-based, stochastic transit assignment models based on MSA. Annals of Operations Research. 144 (1), p.263-285. Kluwer.

Nielsen O, (2000). A Stochastic Transit Assignment Model Considering Differences in Passengers Utility Functions. Transportation Research Part B Methodological, 34 (5), p.377–402. Elsevier Science Ltd.

Prashker J, & Bekhor S, (2004). Route choice models used in the stochastic user equilibrium problem: a review. Transport Reviews, 24 (4), p.437–463.

Prato C, & Bekhor S, (2006). Applying branch & bound technique to route choice set generation. Transportation Research Record, 1985, p.19-28.

Prato C, (2005). Latent factors and route choice behaviour. Ph.D. Thesis, Turin Polytechnic, Italy.

Prato C, (2009). Route choice modeling: past, present and future research directions. Journal of Choice Modelling, 2 (1), p.65–100.

Prato C, (2014). Expanding the applicability of random regret minimization for route choice analysis. Transportation, (41), p.351–375.

Prato C, Rasmussen T, & Otto N, (2014). Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers. Transportation Research Record: Journal of the Transportation Research Board, 2412, p.20–27.

Pravinvongvuth S, & Chen A, (2005). Adaptation of the paired combinatorial logit model to the route choice problem. Transportmetrica, 1 (3), p.223–240.

Ramming S, (2002). Network knowledge and route choice. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA.

Rasmussen T, Nielsen O, Watling D, & Prato C, (2016). The Restricted Stochastic User Equilibrium with Threshold model: Large-scale application and parameter testing. European Journal of Transport Infrastructure Research (EJTIR). 17 (1), p.1-24

Rich J, & Nielsen, O (2015). System convergence in transport models: algorithms efficiency and output uncertainty. European Journal of Transport Infrastructure Research (EJTIR), 15 (3), p.38-62.

Sheffi Y, (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall.

Swait J, & Ben-Akiva M, (1984). Incorporating Random Constraints in Discrete Choice Models: An Application to Mode Choice in Sao Paulo, Brazil. Transportation Research Part B, 21(2), p.103-115.
Vovsha P, (1997). Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. Transportation Research Record 1607, p.6–15.

Watling D, Rasmussen T, Prato C, & Nielsen O, (2018). Stochastic user equilibrium with a bounded choice model. Transportation Research Part B, 114, p.254-280.

Wen C, & Koppelman F, (2001). The generalized nested logit model. Transportation Research Part B, 35 (7), p.627–641.

Xu X, Chen A, Kitthamkesorn S, Yang H, & Lo H.K, (2015). Modeling absolute and relative cost differences in stochastic user equilibrium problem. Transportation Research Part B, 81, p.686-703.

Appendix A: *Discontinuity issue with APSL path size terms when allowing zero choice probabilities*

The APSL path size term for route $i \in R$ is given by:

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{P_i}{\sum_{k \in R} P_k \delta_{a,k}}.$$

The issue is that there are three possible values for $\lim_{\sum_{k \in R} P_k \delta_{a,k} \to 0} \frac{P_i}{\sum_{k \in R} P_k \delta_{a,k}}$:

a) $\lim_{P_i \to 0} \left(\lim_{\sum_{k \in R; k \neq i} P_k \delta_{a,k} \to 0} \frac{P_i}{P_i + \sum_{k \in R; k \neq i} P_k \delta_{a,k}} \right) = 1,$

b) $\lim_{\sum_{k \in R; k \neq i} P_k \delta_{a,k} \to 0} \left(\lim_{P_i \to 0} \frac{P_i}{P_i + \sum_{k \in R; k \neq i} P_k \delta_{a,k}} \right) = 0,$

c) $\lim_{\substack{\sum_{k \in R_m; k \neq i} P_k \delta_{a,k} \to 0 \\ P_i \to 0}} \frac{P_i}{P_i + \sum_{k \in R; k \neq i} P_k \delta_{a,k}} = \frac{1}{\sum_{k \in R} \delta_{a,k}}.$

To demonstrate this, consider the appendix A example network in Fig. A.1 where there are 9 routes.

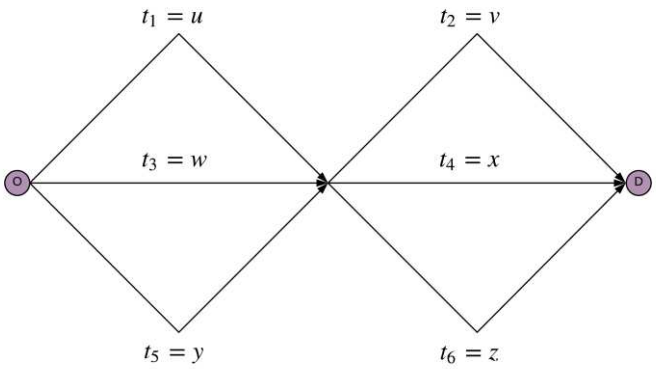


Fig. A.1. Appendix A example network.

Route 1: $1 \to 2$, Route 2: $1 \to 4$, Route 3: $1 \to 6$,
Route 4: $3 \to 2$, Route 5: $3 \to 4$, Route 6: $3 \to 6$,
Route 7: $5 \to 2$, Route 8: $5 \to 4$, Route 9: $5 \to 6$.

Suppose $u = v = w = x = y = z = 1$, and let $P_1 = P_2 = P_3 = \frac{1-P_4}{3}$, $P_4 \in [0,1]$, $P_5 = P_6 = P_7 = P_8 = P_9 = 0$. As $P_4 \to 1$, $P_1 = P_2 = P_3 \to 0$, and the path size terms for Route 1, Route 2, and Route 5 as $P_4 \to 1$ are:

$$\lim_{P_4 \to 1} \gamma_1^{APS}(\boldsymbol{P}) = \lim_{P_4 \to 1} \left(\left(\frac{1}{2}\right) \cdot \left(\frac{P_1}{P_1 + P_2 + P_3} \right) + \left(\frac{1}{2}\right) \cdot \left(\frac{P_1}{P_1 + P_4 + P_7} \right) \right)$$

$$= \left(\frac{1}{2}\right) \cdot \left(\lim_{\substack{P_2 + P_3 \to 0 \\ P_1 \to 0}} \frac{P_1}{P_1 + P_2 + P_3} \right) + \left(\frac{1}{2}\right) \cdot \left(\lim_{\substack{P_4 \to 1 \\ P_1 \to 0}} \left(\lim_{P_7 \to 0} \frac{P_1}{P_1 + P_4 + P_7} \right) \right)$$

$$= \left(\frac{1}{2}\right) \cdot \left(\frac{1}{3}\right) + \left(\frac{1}{2}\right) \cdot (0) = \frac{1}{6}$$

$$\lim_{P_4\to 1}\gamma_2^{APS}(\boldsymbol{P}) = \lim_{P_4\to 1}\left(\left(\frac{1}{2}\right)\cdot\left(\frac{P_2}{P_1+P_2+P_3}\right)+\left(\frac{1}{2}\right)\cdot\left(\frac{P_2}{P_2+P_5+P_8}\right)\right)$$

$$= \left(\frac{1}{2}\right)\cdot\left(\lim_{\substack{P_1+P_3\to 0\\P_2\to 0}}\frac{P_2}{P_1+P_2+P_3}\right)+\left(\frac{1}{2}\right)\cdot\left(\lim_{P_2\to 0}\left(\lim_{P_5+P_8\to 0}\frac{P_2}{P_2+P_5+P_8}\right)\right)$$

$$= \left(\frac{1}{2}\right)\cdot\left(\frac{1}{3}\right)+\left(\frac{1}{2}\right)\cdot(1) = \frac{2}{3}$$

$$\lim_{P_4\to 1}\gamma_5^{APS}(\boldsymbol{P}) = \lim_{P_4\to 1}\left(\left(\frac{1}{2}\right)\cdot\left(\frac{P_5}{P_4+P_5+P_6}\right)+\left(\frac{1}{2}\right)\cdot\left(\frac{P_5}{P_2+P_5+P_8}\right)\right)$$

$$= \left(\frac{1}{2}\right)\cdot\left(\lim_{P_4\to 1}\left(\lim_{\substack{P_5\to 0\\P_6\to 0}}\frac{P_5}{P_4+P_5+P_6}\right)\right)+\left(\frac{1}{2}\right)\cdot\left(\lim_{P_2\to 0}\left(\lim_{\substack{P_5\to 0\\P_8\to 0}}\frac{P_5}{P_2+P_5+P_8}\right)\right)$$

$$= \left(\frac{1}{2}\right)\cdot(0)+\left(\frac{1}{2}\right)\cdot(0) = 0$$

Thus, at $P_4 = 1$ where $P_1 = P_2 = P_3 = P_5 = P_6 = P_7 = P_8 = P_9 = 0$, many cases of $\sum_{k\in R} P_k\delta_{a,k} = 0$ occur, but $\lim_{\sum_{k\in R} P_k\delta_{a,k}\to 0}\frac{P_i}{\sum_{k\in R} P_k\delta_{a,k}}$ either equals 1, 0, or $\frac{1}{3}$, and hence defining the path size terms as either:

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum_{a\in A_i}\frac{t_a}{c_i}\times\begin{cases}\dfrac{P_i}{\sum_{k\in R} P_k\delta_{a,k}} & if\ \sum_{k\in R} P_k\delta_{a,k} > 0\\ 1 & if\ \sum_{k\in R} P_k\delta_{a,k} = 0\end{cases},$$

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum_{a\in A_i}\frac{t_a}{c_i}\times\begin{cases}\dfrac{P_i}{\sum_{k\in R} P_k\delta_{a,k}} & if\ \sum_{k\in R} P_k\delta_{a,k} > 0\\ 0 & if\ \sum_{k\in R} P_k\delta_{a,k} = 0\end{cases},$$

or,

$$\gamma_i^{APS}(\boldsymbol{P}) = \sum_{a\in A_i}\frac{t_a}{c_i}\times\begin{cases}\dfrac{P_i}{\sum_{k\in R} P_k\delta_{a,k}} & if\ \sum_{k\in R} P_k\delta_{a,k} > 0\\ \dfrac{1}{\sum_{k\in R}\delta_{a,k}} & if\ \sum_{k\in R} P_k\delta_{a,k} = 0\end{cases},$$

does not ensure continuity.